GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:22:54 ; Search time 1139.77 Seconds
(without alignments)
8665.841 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1480.4	99.9	1496	2	AAQ28119	Aaq28119 P.falcipa
	2	948.4	64.0	950	2	AAQ28115	Aaq28115 P.falcipa
	3	851.6	57.5	1320	2	AAQ80916	Aaq80916 Plasmodiu
	4	586.4	39.6	1566	2	AAQ05140	Aaq05140 Sequence
	5	503.4	34.0	1371	12	ADO21941	Ado21941 LSA-NRC(H
	6	493.6	33.3	1374	12	ADO21919	Ado21919 LSA-NRC(H
	7	472.8	31.9	988	2	AAQ28117	Aaq28117 P.falcipa
	8	231.8	15.6	1300	12	ADP85917	Adp85917 Synthetic
	9	224.2	15.1	3399	2	AAT05868	Aat05868 Chicken l
	10	183.2	12.4	210	1	AAN80899	Aan80899 Sequence
	11	178.6	12.1	1000	12	ADQ62833	Adq62833 Homopoly-
c	12	178.6	12.1	1000	12	ADQ62832	Adq62832 Homopoly-
	13	177.8	12.0	1686	2	AAQ87587	Aaq87587 DNA encod
	14	175	11.8	3579	3	AAA70099	Aaa70099 Plasmodiu
	15	169	11.4	1998	3	AAA70212	Aaa70212 Plasmodiu
c	16	151.2	10.2	6292	4	AAS46735	Aas46735 Tumour su
c	17	149.2	10.1	6767	4	AAS46608	Aas46608 Tumour su
c	18	144.4	9.7	3662	4	ABL05274	Abl05274 Drosophil
c	19	139.4	9.4	9539	4	AAS45347	Aas45347 Chemicall
c	20	139.4	9.4	9539	6	ABK28180	Abk28180 DNA trans
	21	136.4	9.2	6644	2	AAX33181	Aax33181 Base sequ
	22	136.4	9.2	7372	2	AAX33182	Aax33182 Base sequ
	23	136.4	9.2	7797	2	AAX33180	Aax33180 Cowpox vi
	24	136.4	9.2	7996	2	AAX33184	Aax33184 Base sequ
c	25	136.2	9.2	6668	6	ABL33697	Abl33697 Human imm
c	26	132.6	8.9	5493	13	ADS89675	Ads89675 Oligonucl
c	27	131.2	8.9	778	6	ABQ15588	Abq15588 Oligonucl
	28	131.2	8.9	778	6	ABQ15589	Abq15589 Oligonucl
c	29	129.8	8.8	5493	13	ADS89401	Ads89401 Oligonucl
c	30	129.8	8.8	7442	4	AAS46686	Aas46686 Tumour su
c	31	129	8.7	7597	6	ABL33013	Abl33013 Human imm
	32	128.2	8.7	1039	14	ACL64817	Acl64817 M. xanthu
	33	128.2	8.7	1762	6	ABQ20939	Abq20939 Oligonucl
c	34	128.2	8.7	1762	6	ABQ20938	Abq20938 Oligonucl
c	35	127.6	8.6	1995	6	ABQ20942	Abq20942 Oligonucl
	36	127.6	8.6	1995	6	ABQ20943	Abq20943 Oligonucl
c	37	127.4	8.6	1200	6	ABQ39210	Abq39210 Oligonucl
	38	127.4	8.6	1200	6	ABQ39211	Abq39211 Oligonucl
c	39	127.4	8.6	14006	6	ABL33958	Abl33958 Human imm
c	40	126.8	8.6	3683	8	ABZ10199	Abz10199 Haematopo
c	41	125.4	8.5	3683	8	ABZ10053	Abz10053 Haematopo
c	42	121.8	8.2	3037	6	ABQ54066	Abq54066 Oligonucl
	43	121.8	8.2	3037	6	ABQ54067	Abq54067 Oligonucl
c	44	121.4	8.2	12237	6	ABL34358	Abl34358 Human imm
	45	119.6	8.1	1701	14	ADZ72262	Adz72262 Plasmodiu

ALIGNMENTS

RESULT 1
AAQ28119
ID AAQ28119 standard; DNA; 1496 BP.
XX
AC AAQ28119;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA gene 3' region.
XX
KW Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW paludism; liver stage-specific antigen; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1. .1482
FT /*tag= c
FT /product= "LSA_C-terminal_region"
FT repeat_region 37. .639
FT /*tag= a
FT repeat_unit 37. .87
FT /*tag= b
FT /rpt_type= TANDEM
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.
XX
PR 05-FEB-1991; 91FR-00001286.
XX
PA (INSP) INST PASTEUR.
XX
PI Guerinmarchand C, Druilhe P;
XX
DR WPI; 1992-299985/36.
DR P-PSDB; AAR26944.
XX
PT Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for

```
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 2; Fig 8-10; 81pp; French.
XX
CC   This nucleotide sequence is the 3' part of the P.falciparum liver-stage
CC   specific antigen (LSA) gene. It codes for a polypeptide sequence which
CC   carries a T cell epitope characteristic of a protein produced in
CC   hepatocytes infected with P.falciparum. The polypeptide can be used in
CC   the preparation of vaccines against malaria. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1496 BP; 716 A; 169 C; 300 G; 311 T; 0 U; 0 Other;

  Query Match             99.9%;  Score 1480.4;  DB 2;  Length 1496;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-201;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 2
AAQ28115
ID   AAQ28115 standard; DNA; 950 BP.
XX
AC   AAQ28115;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA-R-NR coding sequence.
XX
KW   Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism; liver stage-specific antigen; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   3. .629
FT                   /*tag=  a
FT                   /rpt_type= TANDEM
FT   repeat_unit     3. .53
FT                   /*tag=  b
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
```

```
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   P-PSDB; AAR26941.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 17; Fig 2; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert (950 bases) encoded LSA-R-NR
CC   containing a 12-repeat region followed by a non-repeat region. Preferred
CC   antigenic polypeptides of the invention are derived from the amino acid
CC   sequence of LSA-R-NR. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 950 BP; 456 A; 127 C; 206 G; 161 T; 0 U; 0 Other;

  Query Match            64.0%;  Score 948.4;  DB 2;  Length 950;
  Best Local Similarity  99.9%;  Pred. No. 6.1e-126;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 3
AAQ80916
ID   AAQ80916 standard; cDNA; 1320 BP.
XX
AC   AAQ80916;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   29-AUG-1995  (first entry)
XX
DE   Plasmodium falciparum LSA-1 repeatless gene cDNA.
XX
KW   Plasmodium falciparum LSA-1 repeatless gene; recombinant poxvirus;
KW   multicomponent multistage malarial vaccines; immunogens;
KW   malaria diagnosis; ss.
XX
OS   Plasmodium falciparum; (pLSARPLS.I4L.1).
XX
PN   WO9428930-A1.
XX
PD   22-DEC-1994.
XX
PF   10-JUN-1994;   94WO-US006652.
XX
PR   11-JUN-1993;   93US-00075783.
PR   09-JUN-1994;   94US-00257073.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Paoletti E,  De Taisne C,  Tine JA;
XX
DR   WPI; 1995-036113/05.
XX
PT   Recombinant poxvirus contg. Plasmodium DNA in non-essential region -
PT   useful in vaccines against malaria and for prodn. of Plasmodium
PT   immunogens.
XX
PS   Claim 3; Fig 11; 183pp; English.
XX
CC   AAQ80916 is the P. falciparum LSA-1 repeatless gene cDNA sequence. New
CC   recombinant poxviruses containing either the SERA, ABRA, Pfhsp70, AMA-1,
CC   Pfs25, Pfs16, CSP, PfSSP2, LSA-1, LSA-1 repeatless, MSA-1, MSA-1 (N-
CC   terminal p83 or C-terminal gp42) genes, or a combination of these in non-
CC   essential regions of their genomes are claimed. These poxviruses (pref.
CC   with a virulence reducing genomic deletion or disruption) can be used as
CC   vaccines against malaria and for the prodn. of Plasmodium immunogens.
CC   These viruses provide multicomponent, multistage vaccines due to their
CC   expression of sporozite, liver stage, blood stage and sexual stage
CC   proteins. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 16-
```

```
CC   OCT-2003 to standardise OS field)
XX
SQ   Sequence 1320 BP; 619 A; 128 C; 227 G; 346 T; 0 U; 0 Other;

  Query Match             57.5%;  Score 851.6;  DB 2;  Length 1320;
  Best Local Similarity   87.8%;  Pred. No. 3.2e-112;
  Matches  929;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        425 ATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 484
              || || || | | |  ||   | ||    ||    |   | | |    | | |   |
Db        263 ATGTAAAAAATGTGTCACAAACAAATTTCAAAAGTCTTTTAAGAAATCTTGGTGTTTCAG 322

Qy        485 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 544
              | | | |     |||||||||      | |  |  || |  | |||    || | | |
Db        323 AGAATATATTCCTTAAAGAAAATAAATTAAATAAGGAAGGGAAATTAATTGAACACATAA 382

Qy        545 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAG 604
              | | ||  |    || || || | |  |     || | || |    |     |||
Db        383 TAAATGATGATGACGATAAAAAAAAATATATTAAAGGGCAAGACGAAAACAGACAAGAAG 442

Qy        605 AAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATT 664
              |        || |   ||    |  |  ||||   ||||||||||||||||||||||||
Db        443 ATCTTGAAGAAAAAGCGCGCGCATCTAAAGAAACGAGGAAGGCTGATACGAAAAAAAATT 502

Qy        665 TAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAA 724
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        503 TAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTAGAAA 562

Qy        725 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 622

Qy        785 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAA 844
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        623 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAA 682

Qy        845 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 742

Qy        905 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 802

Qy        965 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 862

Qy       1025 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 922

Qy       1085 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 982

Qy       1145 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1042

Qy       1205 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1102

Qy       1265 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1162

Qy       1325 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1222

Qy       1385 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1282

Qy       1445 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||
Db       1283 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1320


RESULT 4
AAQ05140
ID   AAQ05140 standard; DNA; 1566 BP.
XX
AC   AAQ05140;
XX
DT   25-MAR-2003  (revised)
DT   05-NOV-1990  (first entry)
XX
DE   Sequence encoding N-terminal of peptide antigen to malarial sporozite.
XX
KW   Malaria; sporozite; vaccine; exoerythrocytic parasites; tetanus toxoid.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             169. .1554
FT                   /*tag=  a
XX
PN   WO9006130-A.
XX
PD   14-JUN-1990.
XX
PF   30-NOV-1988;   88US-00278234.
XX
PR   30-NOV-1988;   88US-00278234.
PR   12-APR-1989;   89US-00337204.
XX
PA   (BIOM-) BIOMEDICAL RES INST.
XX
PI   Hollingdal MR;
XX
DR   WPI; 1990-209624/27.
DR   P-PSDB; AAR05766.
XX
PT   Novel malarial sporozoite antigenic protein - useful as vaccine against
PT   sporozoite(s) and exo:erythrocytic parasites.
XX
PS   Disclosure; Page ?; -pp; English.
XX
CC   Antigen, preferably linked at the C-terminal to a carrier such as tetanus
CC   toxoid, may be used as a vaccine against the malarial sporozite. (Updated
CC   on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 1566 BP; 745 A; 210 C; 330 G; 281 T; 0 U; 0 Other;

  Query Match             39.6%;  Score 586.4;  DB 2;  Length 1566;
  Best Local Similarity   91.7%;  Pred. No. 1.3e-74;
  Matches  620;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||||||||||||||| |||||||||||||| |||||||||||||||||||||||||||
Db        745 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 804

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||||||||||||||||| |||| |||||||||||||| ||||| || |||||||||
Db        805 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 864

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Db        865 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 924

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              || |||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db        925 GATAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 984
```

```
Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||||||||||||||||| |||||||||||||||||| |  || |    ||| ||||
Db        985 GCTAAAGAAAAGTTGCAAGTACAACAAAGCGATTTAGAACGAACTAAAGCATCTACAGAA 1044

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              | ||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db       1045 ACGTTGCAAGAGCGACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 1104

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 1164

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              || ||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db       1165 AGAGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAGCAACAAAGCGATCTA 1224

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||||||||||| |||||||||||||||||||| |||||||| ||||||||||||||||
Db       1225 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAGCAACAAAGAGATTTAGAACAAGAG 1284

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1285 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 1344

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||||||||||||||| || ||||| |||||||||||   ||    || ||    |||
Db       1345 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGTAAAG 1404

Qy        661 AATTTAGAAAGAAAAA 676
                  ||||  | |||
Db       1405 TTGCAAGAACAACAAA 1420
```

RESULT 5
ADO21941
ID ADO21941 standard; DNA; 1371 BP.
XX
AC ADO21941;
XX
DT 12-AUG-2004 (first entry)
XX
DE LSA-NRC(H) construct DNA derived from Malaria parasite LSA-1.
XX
KW LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW malaria parasite P. falciparum; LSA-NRC(H) construct; ds; gene.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT CDS   1. .1371
FT   /*tag= a
FT   /product= "LSA-NRC(H) construct protein derived from
FT   Malaria parasite LSA-1"
XX
PN WO2004044167-A2.
XX
PD 27-MAY-2004.
XX
PF 12-NOV-2003; 2003WO-US036011.
XX
PR 12-NOV-2002; 2002US-0425719P.
XX
PA (REED-) REED ARMY INST RES WALTER.
XX
PI Lanar DE, Hillier CJ, Lyon JA, Angov E, Kumar S, Rogers W;
PI Barbosa A;
XX
DR WPI; 2004-420309/39.
DR P-PSDB; ADO21942.
XX
PT Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT (LSA-1) epitope.
XX
PS Claim 16; SEQ ID NO 25; 90pp; English.
XX
CC The invention relates to a novel recombinant LSA-NRC polypeptide
CC comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC falciparum. The polypeptide of the invention demonstrates antimalarial
CC activities and may be useful as a vaccine during immunotherapy and as a
CC diagnostic agent, preferably for diagnosing malaria. The current sequence
CC is that of the LSA-NRC(H)Mut construct DNA of the invention. The
CC construct is harmonised for expression in Escherichia coli and comprises
CC the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC parasite P. falciparum LSA-1 in addition to a C-terminal His6 tag.
XX
SQ Sequence 1371 BP; 539 A; 358 C; 283 G; 191 T; 0 U; 0 Other;

Query Match 34.0%; Score 503.4; DB 12; Length 1371;
Best Local Similarity 63.1%; Pred. No. 7.8e-63;
Matches 791; Conservative 0; Mismatches 461; Indels 1; Gaps 1;

```
Qy        227 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATA 286
              ||| ||  | |       ||||||  |      || | ||||  |     |||| |  | |
Db         79 AACGAGGAAAAGCATGAGAAGAAACATGTGCTGAGCCACAACTCCTACGAGAAGACTAAA 138

Qy        287 GACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTA 346
              ||      ||| || |  | |     ||  |          ||  | |||  ||||   |
Db        139 AACAACGAAAACAACAAATTCTTTGACAAGGACAAAGAGCTGACGATGAGCAACGTTAAA 198

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA-AAG 405
              ||   |      |  |||   ||| |  |   |  |||    | |||     |||| |
Db        199 AACGTATCCCAGACCAACTTTAAATCCCTCCTGCGCAACCTCGGCGTTTCCGAGAACATC 258

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              || |   || || | |          |  |||  | |||   | || | |   |  | ||
Db        259 TTTCTCAAAGAAAACAAACTGAACAAGGAAGGCAAACTGATTGAACATATCATCAACGAC 318

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
              | |   | |   | || |    | | |     | |||||     || ||| | |
Db        319 GACGATGACAAAAAAAAATACATTAAAGGCCAGGATGAAAATCGCCAGGAAGACCTCGAA 378

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
              ||   ||    ||    |  |  |    | | |     |    ||
Db        379 GAAAAAGCTGCTGAACAGCAGTCGGACCTGGAACAGGAGCGCCTCGCTAAAGAAAAGCTC 438

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
              || ||| | |  |||||||||||| | || ||||| ||  | ||  | |||||  | |||
Db        439 CAGGAGCGCCTCGCTAAAGAAAAGCTCCAGGAGCAACAGCGCGACCTGGAACAGCGCAAG 498

Qy        646 GCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGAT 705
              ||||| ||||||||||  | ||| | ||||||||||| || ||  |  | || |||||
Db        499 GCTGACACGAAAAAAAACCTGGAACGCAAAAAGGAACACGGCGACGTTCTGGCTGAGGAC 558

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATAT 765
              | || || ||  | ||||| |||||||| |||| ||||| ||||| || ||
Db        559 CTGTACGGCCGCCTGGAAATCCCAGCTATCGAACTCCCATCCGAAAACGAACGCGGCTAC 618

Qy        766 TATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825
              || || ||||| ||         | || || || ||  | |||||      | ||     |||
Db        619 TACATCCCACACCAGAGCAGCCTGCCACAAGATAATCGCGGGAACTCCCGCGACAGTAAG 678

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA 885
              |||||    || || |||||||| ||  | |||   ||||| || || || ||||| ||
Db        679 GAAATCAGCATCATCGAAAAAACCAACCGCGAAAGCATTACCACCAACGTGGAAGGCCGC 738

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
              | || || || ||||| || || |||||||||||||| || || || ||||||||||
Db        739 CGCGACATCCACAAAGGCCACCTCGAAGAAAAGAAAGACGGCTCCATCAAACCAGAACAG 798

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              |||||||| |||   ||||| || || || || ||  | ||||| || || |||   ||
```

```
Db        799 AAAGAAGACAAAAGCGCTGATATCCAGAACCACACCCTGGAGACCGTGAACATTAGCGAC 858

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065

Db        859 GTGAACGACTTCCAGATCAGCAAGTACGAGGACGAAATCTCCGCTGAATACGATGACTCC 918

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125

Db        919 CTGATCGACGAAGAAGAAGACGACGAAGATCTGGATGAATTCAAACCAATTGTCCAGTAC 978

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185

Db        979 GATAACTTTCAGGACGAAGAAAATATCGGCATTTACAAAGAACTCGAAGACCTCATCGAG 1038

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245

Db       1039 AAAAACGAAAACCTGGACGACCTGGACGAAGGCATCGAAAAATCCTCCGAAGAACTGAGC 1098

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305

Db       1099 GAAGAAAAAATCAAAAAAGGCAAGAAATACGAAAAAACCAAGGACAACAACTTCAAACCA 1158

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365

Db       1159 AACGACAAATCCCTCTACGACGAGCACATTAAAAAATACAAAAACGACAAGCAAGTGAAC 1218

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425

Db       1219 AAGGAAAAGGAAAAATTTATCAAATCCCTCTTCCACATCTTCGATGGCGATAACGAAATT 1278

Qy       1426 TTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db       1279 CTGCAAATTGTAGACGAACTGAGCGAAGACATCACTAAATACTTCATGAAGCT 1331
```

RESULT 6
ADO21919
ID ADO21919 standard; DNA; 1374 BP.
XX
AC ADO21919;
XX
DT 12-AUG-2004 (first entry)
XX
DE LSA-NRC(H)Mut construct DNA derived from Malaria parasite LSA-1.
XX
KW LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW malaria parasite P. falciparum; LSA-NRC(H)Mut construct; ds; gene;
KW mutant.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1. .1374
FT /*tag= a
FT /product= "LSA-NRC(H)Mut construct protein derived from
FT Malaria parasite LSA-1"
XX
PN WO2004044167-A2.
XX
PD 27-MAY-2004.
XX
PF 12-NOV-2003; 2003WO-US036011.
XX
PR 12-NOV-2002; 2002US-0425719P.
XX
PA (REED-) REED ARMY INST RES WALTER.
XX
PI Lanar DE, Hillier CJ, Lyon JA, Angov E, Kumar S, Rogers W;
PI Barbosa A;
XX
DR WPI; 2004-420309/39.
DR P-PSDB; ADO21920.
XX
PT Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT (LSA-1) epitope.
XX
PS Claim 17; SEQ ID NO 3; 90pp; English.
XX
CC The invention relates to a novel recombinant LSA-NRC polypeptide
CC comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC falciparum. The polypeptide of the invention demonstrates antimalarial
CC activities and may be useful as a vaccine during immunotherapy and as a
CC diagnostic agent, preferably for diagnosing malaria. The current sequence
CC is that of the LSA-NRC(H)Mut construct DNA of the invention. The
CC construct is harmonised for expression in Escherichia coli and comprises
CC the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC parasite P. falciparum LSA-1 in addition to an amino acid insertion
CC within the T5 P. falciparum epitope and a C-terminal His6 tag.
XX
SQ Sequence 1374 BP; 539 A; 358 C; 285 G; 192 T; 0 U; 0 Other;

Query Match 33.3%; Score 493.6; DB 12; Length 1374;
Best Local Similarity 63.1%; Pred. No. 1.9e-61;
Matches 793; Conservative 0; Mismatches 459; Indels 4; Gaps 2;

```
Qy        227 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATA 286

Db         79 AACGAGGAAAAGCATGAGAAGAAACATGTGCTGAGCCACAACTCCTACGAGAAGACTAAA 138

Qy        287 GACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTA 346

Db        139 AACAACGAAAACAACAAATTCTTTGACAAGGACAAAGAGCTGACGATGAGCAACGTTAAA 198

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA-AAG 405

Db        199 AACGTATCCCAGACCAACTTTAAATCCCTCCTGCGCAACCTCGGCGTTTCCGAGAACATC 258

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465

Db        259 TTTCTCAAAGAAAACAAACTGAACAAGGAAGGCAAACTGATTGAACATATCATCAACGAC 318

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525

Db        319 GACGATGACAAAAAAAAAATACATTAAAGGCCAGGATGAAAATCGCCAGGAAGACCTCGAA 378

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585

Db        379 GAAAAAGCTGCTGAACAGCAGTCGGACCTGGAACAGGAGCGCCTCGCTAAAGAAAAGCTC 438

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645

Db        439 CAGGAGCGCCTCGCTAAAGAAAAGCTCCAGGAGCAACAGCGCGACCTGGAACAGCGCAAG 498

Qy        646 GCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGAT 705

Db        499 GCTGACACGAAAAAAAACCTGGAACGCAAAAAGGAACACGGCGACGTTCTGGCTGAGGAC 558

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATAT 765

Db        559 CTGTACGGCCGCCTGGAAATCCCAGCTATCGAACTCCCATCCGAAAACGAACGCGGCTAC 618

Qy        766 TATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825

Db        619 TACATCCCACACCAGAGCAGCCTGCCACAAGATAATCGCGGGAACTCCCGCGACAGTAAG 678

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA 885

Db        679 GAAATCAGCATCATCGAAAAAACCAACCGCGAAAGCATTACCACCAACGTGGAAGGCCGC 738

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945

Db        739 CGCGACATCCACAAAGGCCACCTCGAAGAAAAGAAAGACGGCTCCATCAAACCAGAACAG 798
```

```
Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              |||||||| |||   ||||| || || || || ||  | ||||| || || |||   ||
Db        799 AAAGAAGACAAAAGCGCTGATATCCAGAACCACACCCTGGAGACCGTGAACATTAGCGAC 858

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
              || || || || || || || ||||| ||||| |||||     |||||||| || || ||
Db        859 GTGAACGACTTCCAGATCAGCAAGTACGAGGACGAAATCTCCGCTGAATACGATGACTCC 918

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              | || || |||||||||||| || |||||  | || ||||| || || ||||| || ||
Db        919 CTGATCGACGAAGAAGAAGACGACGAAGATCTGGATGAATTCAAACCAATTGTCCAGTAC 978

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
              || || || || || |||||||| || || ||||| |||||||| |||||  | || |||
Db        979 GATAACTTTCAGGACGAAGAAAATATCGGCATTTACAAAGAACTCGAAGACCTCATCGAG 1038

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              ||||| |||||  | || ||  | || ||||| || |||||||| || |||||| |
Db       1039 AAAAACGAAAACCTGGACGACCTGGACGAAGGCATCGAAAAATCCTCCGAAGAACTGAGC 1098

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              |||||||||| |||||||| |||||||| ||||| || || || ||||||
Db       1099 GAAGAAAAAATCAAAAAAGGCAAGAAATACGAAAAAACCAAGGACAACAACTTCAAACCA 1158

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
              || || |||    | || || ||||| |||||||||| ||||| || ||||| || ||
Db       1159 AACGACAAATCCCTCTACGACGAGCACATTAAAAAATACAAAAACGACAAGCAAGTGAAC 1218

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
              |||||||||||||||| || |||||  | || || || || || || || || ||||||
Db       1219 AAGGAAAAGGAAAAATTTATCAAATCCCTCTTCCACATCTTCGATGGCGATAACGAAATT 1278

Qy       1426 TTACAGATCGTGGATGA---GTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478
              | || || || || ||   |||    ||||| || |||||||| || ||||| ||
Db       1279 CTGCAAATTGTAGACGAACGGTTGAGCGAAGACATCACTAAATACTTCATGAAGCT 1334
```

RESULT 7
AAQ28117
ID AAQ28117 standard; DNA; 988 BP.
XX
AC AAQ28117;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA gene 5' region.
XX
KW Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW paludism; liver stage-specific antigen; ss.
XX
OS Plasmodium falciparum.
XX
FH Key   Location/Qualifiers
FT 5'UTR   1. .32
FT   /*tag= a
FT misc_feature   33. .988
FT   /*tag= d
FT   /note= "5' part of LSA gene"
FT repeat_region   492. .988
FT   /*tag= b
FT repeat_unit   492. .542
FT   /*tag= c
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.
XX
PR 05-FEB-1991; 91FR-00001286.
XX
PA (INSP) INST PASTEUR.
XX
PI Guerinmarchand C, Druilhe P;
XX
DR WPI; 1992-299985/36.
DR P-PSDB; AAR26943.
XX
PT Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT vaccination against, treatment of and diagnosis of malaria.
XX
PS Claim 11; Fig 6; 81pp; French.
XX
CC This nucleotide sequence is the 5' part of the P.falciparum liver-stage
CC specific antigen (LSA) gene. It codes for a polypeptide sequence which
CC carries a T cell epitope characteristic of a protein produced in
CC hepatocytes infected with P.falciparum. The polypeptide can be used in
CC the preparation of vaccines against malaria. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 988 BP; 465 A; 120 C; 184 G; 219 T; 0 U; 0 Other;

  Query Match   31.9%; Score 472.8; DB 2; Length 988;
  Best Local Similarity 95.7%; Pred. No. 1.7e-58;
  Matches 486; Conservative   0; Mismatches   22; Indels   0; Gaps   0;

```
Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |  |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db        541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db        661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

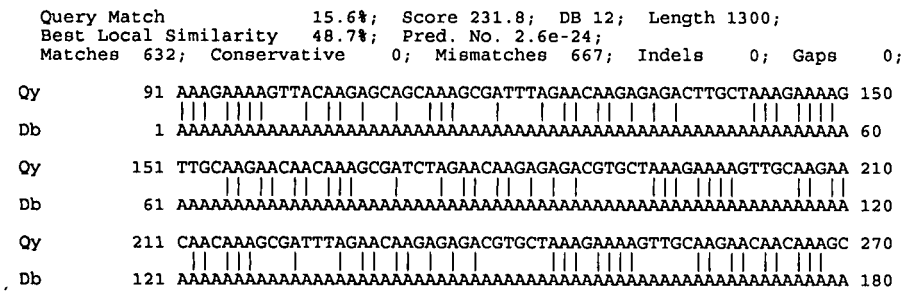
Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

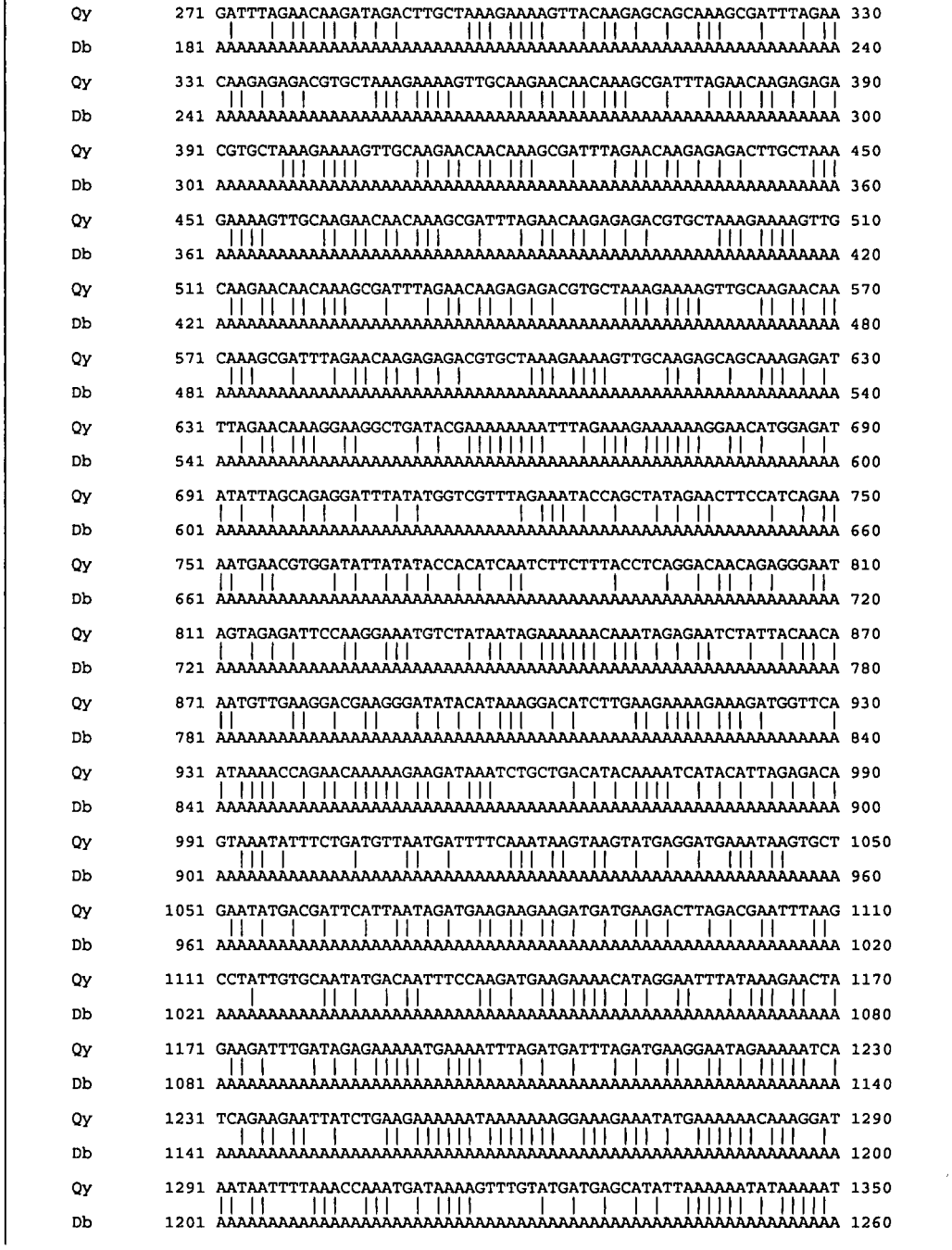
Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988
```

RESULT 8
ADP85917
ID ADP85917 standard; DNA; 1300 BP.
XX
AC ADP85917;
XX
DT 26-AUG-2004 (first entry)
XX
DE Synthetic construct #1.
XX
KW Nanopore data analysis ; polymer ; polynucleotide;

KW single-molecule resolution; ds.
XX
OS Synthetic.
XX
PN US2004110205-A1.
XX
PD 10-JUN-2004.
XX
PF 23-SEP-2003; 2003US-00668749.
XX
PR 23-SEP-2002; 2002US-0412959P.
XX
PA (WANG/) WANG H.
XX
PI Wang H;
XX
DR WPI; 2004-467656/44.
XX
PT Nanopore analysis system for analyzing polynucleotides includes nanopore
PT device having structure with aperture, and nanopore data analysis system
PT that can generate nanopore data points corresponding to each target and
PT non-target polymer.
XX
PS Disclosure; SEQ ID NO 1; 17pp; English.
XX
CC The invention relates to a nanopore analysis system that comprises a
CC nanopore device and a nanopore data analysis system. The nanopore device
CC has a structure having an aperture. The nanopore data analysis system can
CC generate nanopore data points corresponding to each target polymer and
CC each non-target polymer traversing the aperture; form a distribution
CC pattern of the data points; and analyse a distribution of target polymer
CC data points in the distribution pattern. The distribution pattern
CC includes data cluster(s). The nanopore data analysis system is stored on
CC a computer-readable medium. A mechanism is provided for analysing the
CC distribution of target polynucleotide data points in the distribution
CC pattern. The nanopore data analysis system can also analyse the
CC distribution of the non-target polynucleotide data points. It can also
CC determine a ratio between the target polynucleotide data points and the
CC non-target polynucleotide data points. It can further analyse the
CC distribution of target polynucleotide data points between the two data
CC clusters; compare the distribution of the target polynucleotide data
CC points between the two data clusters to a phosphorylation state standard
CC distribution; and determine a ratio of phosphorylated target
CC polynucleotide to non-phosphorylated target polynucleotides. It can also
CC determine a cluster score for the target polynucleotide data points in a
CC defined area; and compare the cluster score for the target polynucleotide
CC data points to a cluster score for a chemical integrity standard density
CC distribution for the defined area in a distribution of a target
CC polynucleotide standard. The invention is useful for analysing polymers,
CC such as polynucleotides. The inventive nanopore analysis system
CC potentially provides high speed sampling with single-molecule resolution,
CC which may enable unprecedented dynamic range and sensitivity in analysis
CC of samples containing charged polymers such as polynucleotides or
CC polypeptides. The present sequence is a synthetic construct used in the
CC invention.
XX
SQ Sequence 1300 BP; 1300 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

Query Match 15.6%; Score 231.8; DB 12; Length 1300;
Best Local Similarity 48.7%; Pred. No. 2.6e-24;
Matches 632; Conservative 0; Mismatches 667; Indels 0; Gaps 0;





```
Qy       1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db       1261 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1299


RESULT 9
AAT05868
ID   AAT05868 standard; DNA; 3399 BP.
XX
AC   AAT05868;
XX
DT   27-AUG-2003  (revised)
DT   14-AUG-1996  (first entry)
XX
DE   Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
KW   Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW   immunisation; vaccination; ss.
XX
OS   Leucocytozoon sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .3399
FT                   /*tag= a
FT   misc_feature    1150. .3218
FT                   /*tag= b
FT                   /note= "fragment referred to in the claims, for use as
FT                   insert in a recombinant vaccine against chicken
FT                   leucocytozoan disease"
XX
PN   JP07284392-A.
XX
PD   31-OCT-1995.
XX
PF   19-APR-1994;   94JP-00080643.
XX
PR   19-APR-1994;   94JP-00080643.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (KITA ) KITASATO KENKYUSHO SH.
XX
DR   WPI; 1996-006311/01.
DR   P-PSDB; AAR97866.
XX
PT   Chicken leucocytozoan immunogenic protein - used in a recombinant vaccine
PT   against chicken leucocytozoan disease.
XX
PS   Claim 6; Page 6-9; 35pp; Japanese.
XX
CC   AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or
CC   a fragment of it can be used in a recombinant vaccine to immunise against
CC   chicken leucocytozoan disease. The DNA is used in a vector and
CC   operatively linked to an expression regulatory sequence as in standard
CC   practice. (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 U; 0 Other;

  Query Match           15.1%;  Score 224.2;  DB 2;  Length 3399;
  Best Local Similarity   48.6%;  Pred. No. 2.9e-23;
  Matches  712;  Conservative    0;  Mismatches  738;  Indels   15;  Gaps    3;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db       1685 AAGTAACACATGAAGAAATAGAAAAAGAAGAGCATGAAGAAGTAATACATGAAGAAGAAA 1744

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db       1745 AAGAAGAAGTAACACATGAAGAAATAGAAAAAGAAGAGCATGAAGAAGTAATACATGAAG 1804

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db       1805 AAGAAAAAGAAGAAGTAACACATGAAG------AAAAAGAAAAAGAAGAGCATGAAGAAG 1858

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db       1859 TAATACATGAAGAAGAAAAAGAAGAAGTAACACATGAAGAAATAGAAAAAGAAGAGCATG 1918

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db       1919 AAGAAGTAATACATGAAGAAGAAAAAGAAGAAGTAACACATGAAGAAATAGAAAAAGAAG 1978

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db       1979 AGCATGAAGAAGTAATACATGAGGAAGAAAAAGAAGAAGTAACACATGAAGAAATAGAAA 2038

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db       2039 AAGAA------GAGCATGAAGAAGTAATACATGAAGAAGAAAAAGAAGAAGTAACACATG 2092

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db       2093 AAGAAATAGAAAAAGAAGAGCATGAAGAAGTAATACATGAAGAAGAAAAAGAAGAAGTAA 2152

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db       2153 CACATGAAGAAAAAGAAAAAGAAGAGCATGAAGAAGTAATACATGAAGAAGAAAAAGAAG 2212

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db       2213 AAGTAACACATGAAGAAGAAGAAGAAAAAGTAACACATGAAGAAGAAGAAGAAAAAGTAA 2272

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db       2273 TACATGAAGAAGAAAAAGAAGAAGTAATACATGAAGAAGAAAAAGAAGAAGTAACACATG 2332

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db       2333 AAGAAGAAAAAGAAGAAGTAACACATGAAGAAGAAAAAGAAGAAGTAACACATGAAGAAG 2392

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db       2393 AAAAAGAAGAAGTAACACATGAAGAAGAAAAAGAAGAAGTAACACATGAAGAAGAAGAAA 2452

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
Db       2453 AAGTAACACATGAAGAAGAAAAAGAAGAAGTAACACATGAAGAAGAAAAAGAAGAAGTAA 2512

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db       2513 CACATGAAGAAGAAGAAAAAGTAACACATGAAGAAGAAGAAAAAGTAACACATGAAGAAG 2572

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db       2573 AAGAAAAAGTAACATATGAAGAAGAAGAAGAAGAAGAAAAAGTAACACATGAAGAAG 2632

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       2633 AAGAAAAAGTAACACATGAAGAAGAAGAAAAAGTAACACATGAAGAAGAAGAAAAAGTAA 2692

Qy       1022 TAAGTAAGTATGAGG---ATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAG 1078
Db       2693 TACATGAAGAAGAAGAAAAAGAAGAAGATGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAG 2752

Qy       1079 AAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAG 1138
Db       2753 AGGAAGAAGAAGAAGAAGAAGATGAGGAAGAGGAAGAAGAAGAAGAAGATGAGGAAG 2812

Qy       1139 ATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATT 1198
Db       2813 AGGAAGAAGAAGAAGAAAATGAGGAAGAAGAAGAAGAAGAAAATAAGGAAGAAGAAGAAG 2872

Qy       1199 TAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAA 1258
Db       2873 AAGAAAAAGAAGAGCATGAAGAAGAAGTAACACATGAAGAAGAAGAAGAAAAAGTAACAC 2932

Qy       1259 AAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTT 1318
```

```
              |  ||| | |  || | | || |    | ||  |  |   |  | || ||||  ||
Db       2933 ATGAAGAAGAAGAAAAAGTAACACATGAAGAAGAAGAAAATGTAACATATGAAGAAGAAG 2992

Qy       1319 TGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAA 1378
                | | |        | || || |  |||| |  |  || |   |  | ||||| | ||
Db       2993 AAGAAAAAGTAACACATGAAGAAGAAGAAAAAGTAACACATGAAGAAGAAGAAAAAGTAA 3052

Qy       1379 AATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGG 1438
                |    | ||  |    |   | || ||| | |||  | ||||   ||    |  |
Db       3053 CACATGAAGAAGAAGAAAATGTAACATATGAAGAAGAAGAAGAAAAAGTAACACATGAAG 3112

Qy       1439 ATGAGTTATCTGAAGATATAACTAA 1463
                | ||   |    |||  || |  ||
Db       3113 AAGAAGAAGAAAAAGTAATGAAGAA 3137


RESULT 10
AAN80899
ID   AAN80899 standard; DNA; 210 BP.
XX
AC   AAN80899;
XX
DT   14-MAY-2003  (revised)
DT   25-MAR-2003  (revised)
DT   19-SEP-1990  (first entry)
XX
DE   Sequence encoding an epitope characteristic of a protein produced in
DE   liver cells infected with Plasmodium falciparum.
XX
KW   Plasmodium falciparum; vaccine; malaria diagnosis;
KW   malaria antibody detection.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .204
FT                   /*tag=  a
XX
PN   WO8805785-A.
XX
PD   11-AUG-1988.
XX
PF   09-FEB-1988;   88WO-FR000074.
XX
PR   09-FEB-1987;   87FR-00001543.
XX
PA   (INSP ) INST PASTEUR.
PA   (MARC/) MARCHAND C.
PA   (CNRS ) CENT NAT RECH SCI.
XX
PI   Druilhe P,  Puijalonme O,  Marchand C,  Langsley G;
XX
DR   WPI; 1988-235148/33.
DR   P-PSDB; AAP81355.
XX
PT   New polypeptide(s) contg. plasmodium falciparum epitope - useful for
PT   malaria vaccination or diagnosis.
XX
PS   Example 1; Fig 1; 37pp; French.
XX
CC   This sequence is contained in clone DG307 of a genomic DNA library of P.
CC   falciparum. AAN80899 is composed entirely of a repeating motif of 51 base
CC   pairs. The type of repitition coded by DG307 seems to be conserved which
CC   implies that it may be present in all P. falciparum strains. The AA
CC   repeat unit is the basis of the peptides claimed in the patent, which are
CC   useful for the prodn. of of malaria vaccines and as immunoassay reagents
CC   for diagnosis of malaria by detection of antibodies. (Updated on 25-MAR-
CC   2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)
CC   (Updated on 14-MAY-2003 to correct PS field.)
XX
SQ   Sequence 210 BP; 97 A; 33 C; 47 G; 33 T; 0 U; 0 Other;

  Query Match             12.4%;  Score 183.2;  DB 1;  Length 210;
  Best Local Similarity   95.9%;  Pred. No. 2.3e-17;
  Matches  188;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         83 GACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTA 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 GACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTA 67

Qy        143 AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGT 202
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db         68 AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACTAGCTAAAGAAAAGT 127

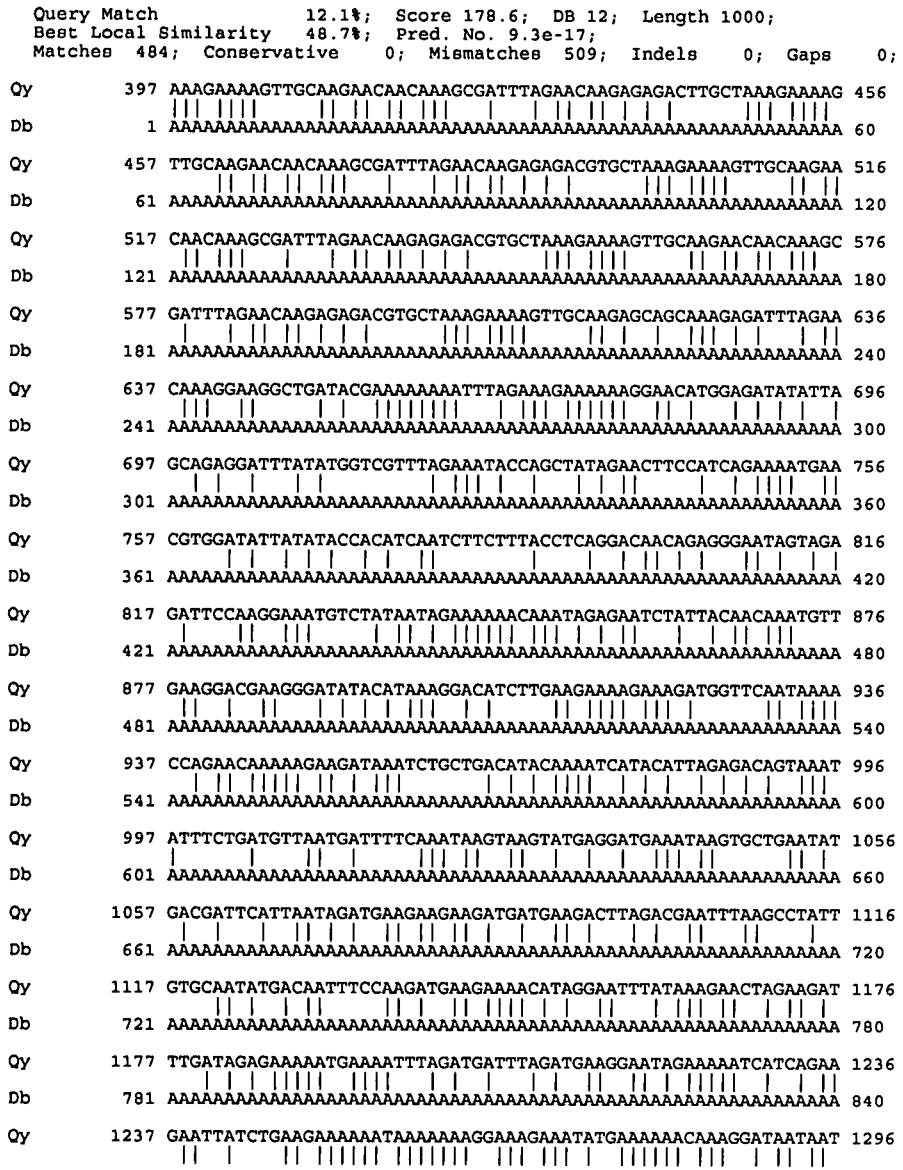
Qy        203 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAAC 262
              | || |  |||||||||||| |||||||||||||| ||||||||||||||||||||||
Db        128 TACAGGGGCAACAAAGCGATCTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAAC 187

Qy        263 AACAAAGCGATTTAGA 278
              ||||||||||||||||
Db        188 AACAAAGCGATTTAGA 203


RESULT 11
ADQ62833
ID   ADQ62833 standard; RNA; 1000 BP.
XX
AC   ADQ62833;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Homopoly-A contaminant for RNaseH activity assay.
XX
KW   ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW   fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW   reverse transcriptase.
XX
OS   Synthetic.
XX
PN   WO2004059012-A1.
XX
PD   15-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US040879.
XX
PR   23-DEC-2002; 2002US-0436125P.
XX
PA   (AMHP ) WYETH.
XX
PI   Olson MW,  O'connell JF;
XX
DR   WPI; 2004-543471/52.
XX
PT   Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT   for detecting and monitoring RNase H activity, comprises hybridizing a
PT   target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS   Example 1; SEQ ID NO 4; 61pp; English.
XX
CC   The invention relates to a method of detecting a nuclease-mediated
CC   cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC   agent by hybridizing a target nucleic acid to a fluorescently labelled
CC   oligonucleotide probe complementary to the target nucleic acid and
CC   containing a fluorophore at one terminus and a quenching group at the
CC   other terminus and contacting the probe-target hybrid with an agent
CC   having nuclease activity. When the oligonucleotide probe is unhybridised
CC   to the target nucleic acid, the probe adopts a conformation that places
CC   the fluorophore and quencher in such proximity that the quencher quenches
CC   the fluorescent signal of the fluorophore. Formation of the probe-target
CC   hybrid causes sufficient separation of the fluorophore and quencher to
CC   reduce quenching of the fluorescent signal of the fluorophore followed by
CC   contacting the probe-target hybrid with an agent having nuclease activity
CC   to selectively cleave the target nucleic acid and thus release the intact
CC   probe. Detection of the release of the probe is by measuring a decrease
CC   in the fluorescent signal of the fluorophore as compared to the signal of
CC   the probe-target hybrid. The methods are useful for detecting and
```

CC monitoring RNaseH activity in real time. The methods are also useful for
CC monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC the invention, the activity of the RNaseH from E. coli or HIV reverse
CC transcriptase is measure by the method of the invention. in order to test
CC the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC or RNA contaminant is added to the assay reaction. This sequence
CC corresponds to the single stranded homopolymeric polyA RNA contaminant
CC for testing this activity.
XX
SQ Sequence 1000 BP; 1000 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

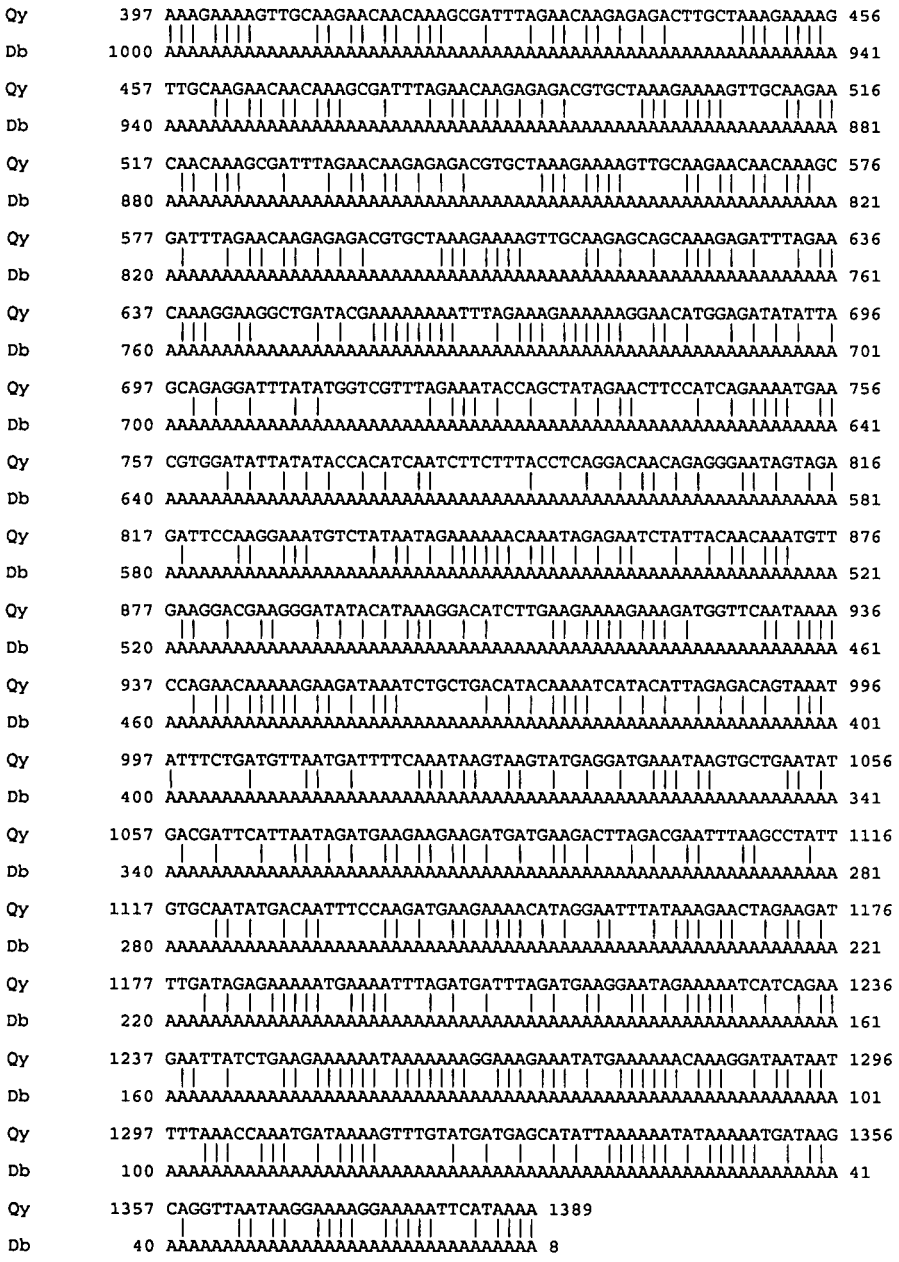
Query Match 12.1%; Score 178.6; DB 12; Length 1000;
Best Local Similarity 48.7%; Pred. No. 9.3e-17;
Matches 484; Conservative 0; Mismatches 509; Indels 0; Gaps 0;



RESULT 12
ADQ62832/c
ID ADQ62832 standard; RNA; 1000 BP.
XX
AC ADQ62832;
XX
DT 07-OCT-2004 (first entry)
XX
DE Homopoly-U contaminant for RNaseH activity assay.
XX
KW ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW reverse transcriptase.
XX
OS Synthetic.
XX
PN WO2004059012-A1.
XX
PD 15-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US040879.
XX
PR 23-DEC-2002; 2002US-0436125P.
XX
PA (AMHP) WYETH.
XX
PI Olson MW, O'connell JF;
XX
DR WPI; 2004-543471/52.
XX
PT Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT for detecting and monitoring RNase H activity, comprises hybridizing a
PT target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS Example 1; SEQ ID NO 3; 61pp; English.
XX
CC The invention relates to a method of detecting a nuclease-mediated
CC cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC agent by hybridizing a target nucleic acid to a fluorescently labelled
CC oligonucleotide probe complementary to the target nucleic acid and
CC containing a fluorophore at one terminus and a quenching group at the
CC other terminus and contacting the probe-target hybrid with an agent
CC having nuclease activity. When the oligonucleotide probe is unhybridised
CC to the target nucleic acid, the probe adopts a conformation that places
CC the fluorophore and quencher in such proximity that the quencher quenches
CC the fluorescent signal of the fluorophore. Formation of the probe-target
CC hybrid causes sufficient separation of the fluorophore and quencher to
CC reduce quenching of the fluorescent signal of the fluorophore followed by
CC contacting the probe-target hybrid with an agent having nuclease activity
CC to selectively cleave the target nucleic acid and thus release the intact
CC probe. Detection of the release of the probe is by measuring a decrease
CC in the fluorescent signal of the fluorophore as compared to the signal of
CC the probe-target hybrid. The methods are useful for detecting and
CC monitoring RNaseH activity in real time. The methods are also useful for
CC monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC the invention, the activity of the RNaseH from E. coli or HIV reverse
CC transcriptase is measure by the method of the invention. in order to test
CC the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC or RNA contaminant is added to the assay reaction. This sequence
CC corresponds to the single stranded homopolymeric polyU RNA contaminant
CC for testing this activity.
XX

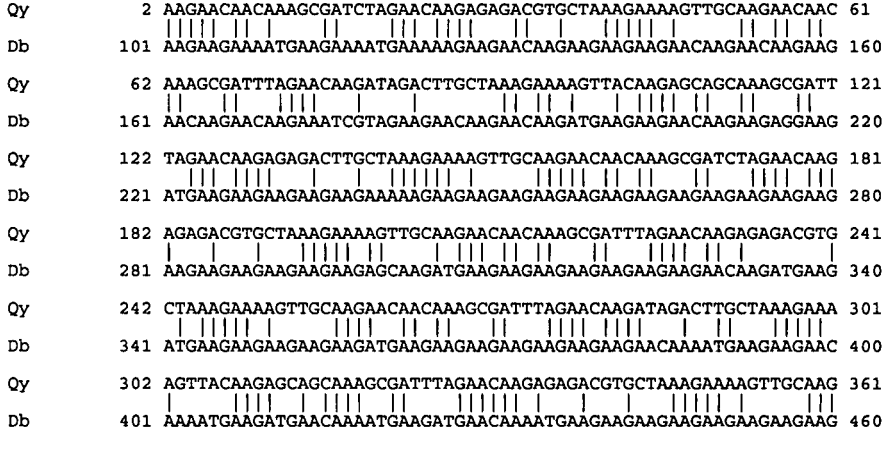
SQ Sequence 1000 BP; 0 A; 0 C; 0 G; 0 T; 1000 U; 0 Other;

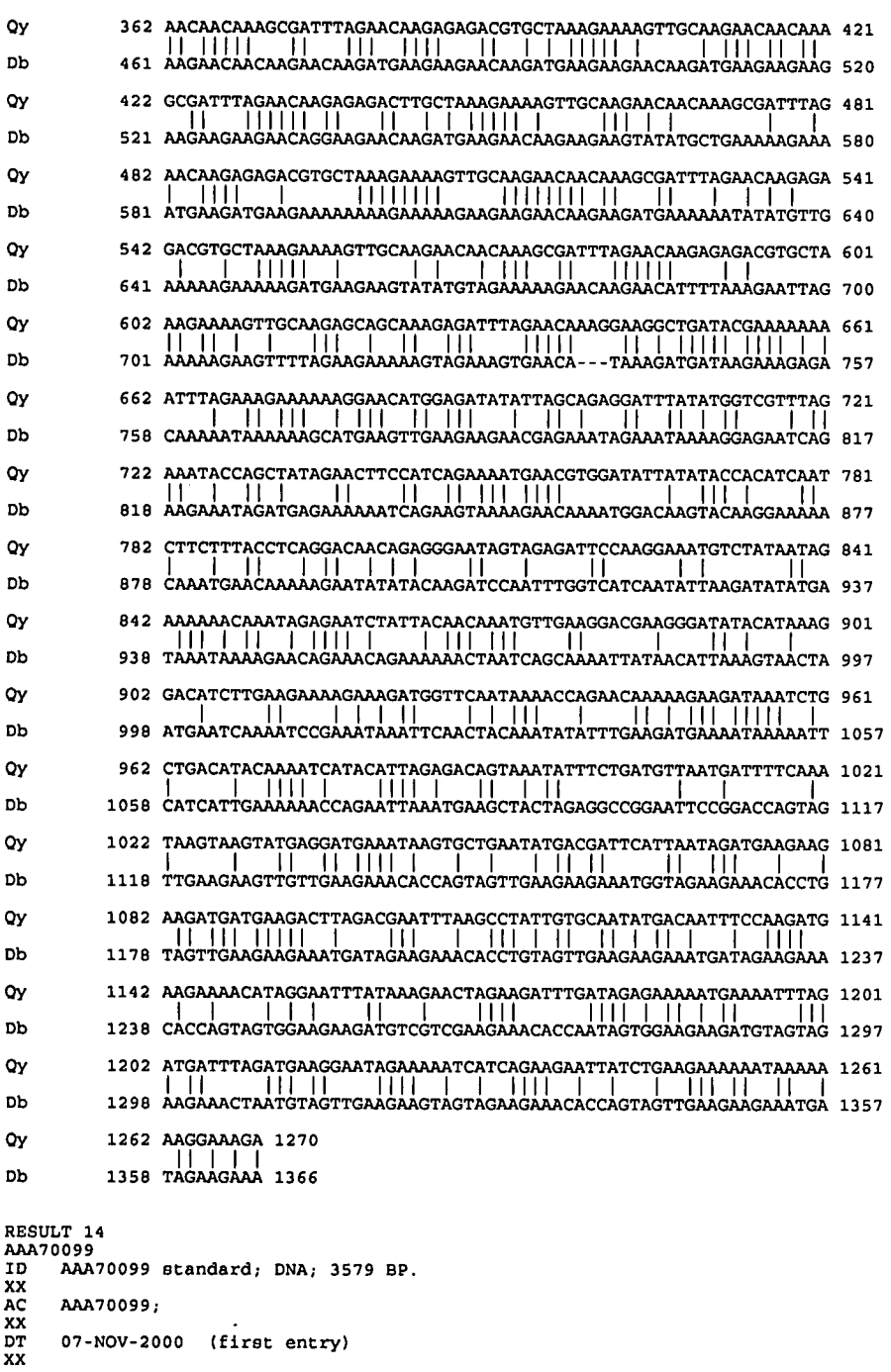
Query Match 12.1%; Score 178.6; DB 12; Length 1000;
Best Local Similarity 48.7%; Pred. No. 9.3e-17;
Matches 484; Conservative 0; Mismatches 509; Indels 0; Gaps 0;



RESULT 13
AAQ87587
ID AAQ87587 standard; DNA; 1686 BP.
XX
AC AAQ87587;
XX
DT 27-AUG-2003 (revised)
DT 19-DEC-1995 (first entry)
XX
DE DNA encoding Leucocytozoan protozoa structural protein epitope.
XX
KW leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
KW leucocytozoanosis; treatment; ss.
XX
OS Leucocytozoon.
XX
PN JP07089995-A.
XX
PD 04-APR-1995.
XX
PF 10-SEP-1993; 93JP-00226078.
XX
PR 10-SEP-1993; 93JP-00226078.
XX
PA (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA (NISS-) NISSEIKEN KK.
XX
DR WPI; 1995-167252/22.
DR P-PSDB; AAR70491.
XX
PT Immune inducing polypeptide against Leucocytozoan protozoa - useful in
PT production of vaccines for treatment of leucocytozoanosis in fowl.
XX
PS Claim 1; Page 12-14; 20pp; Japanese.
XX
CC AAQ87587-89 encode polypeptides having a whole or partial epitope of a
CC structural protein of Leucocytozoan protozoa (see AAR70491-93). The
CC polypeptides and DNA encoding them are useful in the production of
CC vaccines for the treatment of leucocytozoanosis of fowl. (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 U; 0 Other;

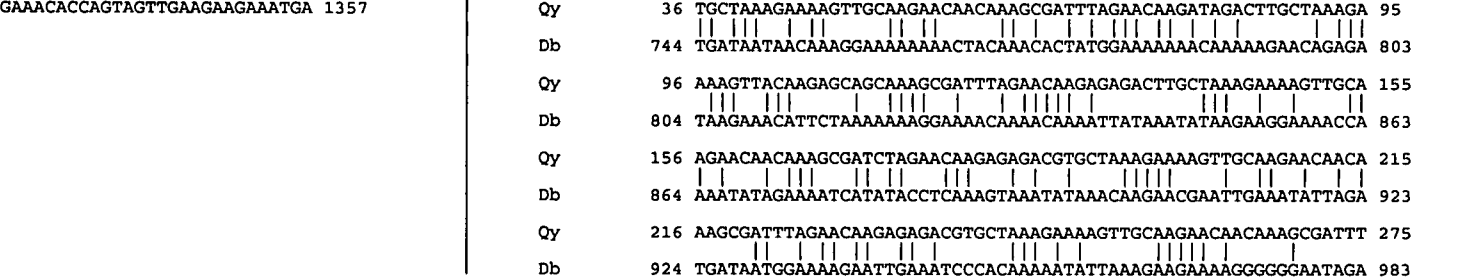
Query Match 12.0%; Score 177.8; DB 2; Length 1686;
Best Local Similarity 46.8%; Pred. No. 1.2e-16;
Matches 594; Conservative 0; Mismatches 672; Indels 3; Gaps 1;





RESULT 14
AAA70099
ID AAA70099 standard; DNA; 3579 BP.
XX
AC AAA70099;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:232.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US026796.
XX
PR 05-NOV-1998; 98US-0107131P.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 457-458; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 3579 BP; 1904 A; 398 C; 552 G; 725 T; 0 U; 0 Other;

Query Match 11.8%; Score 175; DB 3; Length 3579;
Best Local Similarity 46.2%; Pred. No. 2.8e-16;
Matches 660; Conservative 0; Mismatches 760; Indels 9; Gaps 2;



```
Qy        276 AGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGA 335
              | ||   ||||     |  || || | |  || ||     |  ||   ||   ||||  | |
Db        984 AAAAACTGATACTACTAATATAGCAGATATAAAAATAAAAAAAGAAGAAAGAGAAACAAA 1043

Qy        336 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 395
                  || ||||  | ||| | | |  ||  ||| || ||  |   | |  ||
Db       1044 AGATGAAAAAGAAAAAAACATTCAACAGCTAGTAAAAGATGTACAATTAATAAAAGTTGG 1103

Qy        396 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAA 455
              |||||| |    | ||  || ||   ||   ||||  |         ||   |||| |
Db       1104 AGAAGAAACGAAAGACGACGAAAAAGAAGACAAAGAAGGAACAGATGATGAAGAAGATAC 1163

Qy        456 GTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 515
              | ||| || | |  |||   ||| |||| | |  || | ||||| |        ||
Db       1164 AGATGATGAAGAAGATACAGATGATGAAGAAGATACAGATGATGAAGAAGATACAAGTGA 1223

Qy        516 ACAACAAA------GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 569
              || |||       | |||  |||| |  |    |      |  ||| | |     || ||
Db       1224 TGAAGAAACTACAGGTGATCAAGAAAACAAAGAAGAAACAGAAGTAGACGAAAAAAAAAC 1283

Qy        570 ACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGA 629
              | |||  |    |||| ||      |   |  |||||| | |   || | |  ||  |||
Db       1284 AGAAAAAGCCGAAGAAGAATTAGAAGAAGACAAAGAAGAATCAGAAAAAGACAAAGAAGA 1343

Qy        630 TTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGA 689
              | |||| ||   || |  ||  |  |||| ||  ||||  |  | |  || |   |||
Db       1344 ATCAGAAAAAGATAAAGAAGAATCAGAAAAAGATAAAGAAGAATCAGAAAAAGACAAAGA 1403

Qy        690 TATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGA 749
              | |   | ||| |||  | |      |  |||   |  ||  |  ||| |      | |
Db       1404 AAAAACTGAAGAAGATGAAGAAAAAACTGAAGACGAAAAAGGAACGGAAGTATACAAAAA 1463

Qy        750 AAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAA 809
              | |   |  || | || | | |  | | |        |      |||  |      |   |
Db       1464 AGAAACAGATGTAGATGAAAAAAAAGAAAAGGGAGAATATGGAGAGGGAACAGATGATGA 1523

Qy        810 TAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAAC 869
                  |    || |||             |||| || |     ||||   |  |||
Db       1524 AGAAGACAAAGAAAAAGAAGAAGACGACGAAGAAACAAAAGTAGAAGAAAAGAAAACAGA 1583

Qy        870 AAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTC 929
              ||| |  ||||   |||  |||    |  |||      |||   | | |||||||
Db       1584 AAAAGACGAAGAGGGAACAGATTATGAAGAAGATACAGATGATTCAGACAAAGATGAAGA 1643

Qy        930 AATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGAC 989
              || ||||  |||| ||||  ||     ||   |  ||   | ||| | |   |   || |
Db       1644 AACAAAAGTAGAAGAAAAGAAAACAGAAAGAGACGAAGAAGAAACTGAAGAAGACGAAAA 1703

Qy        990 AGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGC 1049
              || ||         ||    ||  |     ||| |     |  | |     ||       |
Db       1704 AGAAACAGAAGTAGAAAAAAAGAAAACAGAAAAAGACGAAGAGGGAACAGATTATGAAGA 1763

Qy       1050 TGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAA 1109
               ||    || ||||||   | ||||| ||||  ||| |  |||||   ||||| |    |
Db       1764 AGATACAGATGATTCAGACAAAGATGTAGAAACAGAAGTAGAAGAAACAGACGCAGAAGA 1823

Qy       1110 GCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACT 1169
                  |    | ||             | |||||||| ||| || |     | | ||| ||
Db       1824 CAAAGAAGAAAACGAAGAGGGAACAGATGATGAAGAAGACAAAGTAGAAGAAACAGACCT 1883

Qy       1170 AGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATC 1229
              ||| ||      |||  |    ||| |      | ||    ||  ||| || ||| |||
Db       1884 AGACGACCAAGAAGAAGACGGAGAAGAAGATAAAGAAGACGACAAAGAAAAAGACAAAGA 1943

Qy       1230 ATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGA 1289
              |   ||  ||  |  |||  || ||| | | |||| | |   |  | ||||||   || ||
Db       1944 AGATGACAAAGAAGATGACAAAGAAAAAGACAAAGAAGATGACAAAGAAAAATACAAAGA 2003

Qy       1290 TAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAA 1349
              ||  |    |||  | |||| |||        |  | ||  |    ||| || |||| ||
Db       2004 AGATGAC---AAAGAAGATGACAAAGAAGATGACAAAGAAAAAGACAAAGAAGATAACAA 2060
```

```
Qy       1350 TGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGA 1409
                 || ||  |      |  |  | || ||||||  || | ||           | | |    |
Db       2061 AGAAAAAGACAAAGAAGATAACAAAGAAAAAGACAAAGAAGATGACAAAGAAAAAGACAA 2120

Qy       1410 CGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
                 | |||  |  ||     | |  |  |  ||| |    |  |||| | |
Db       2121 AGAAGATGACAAAGAAAAAGACAAAGAAGATAACAAAGAAAAAGACAAA 2169
```

RESULT 15
AAA70212
ID AAA70212 standard; DNA; 1998 BP.
XX
AC AAA70212;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:345.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US026796.
XX
PR 05-NOV-1998; 98US-0107131P.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 537; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 1998 BP; 1127 A; 141 C; 338 G; 392 T; 0 U; 0 Other;

Query Match 11.4%; Score 169; DB 3; Length 1998;
Best Local Similarity 45.6%; Pred. No. 2e-15;

Matches 593; Conservative 0; Mismatches 707; Indels 0; Gaps 0;

```
Qy        80 ATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTG 139
Db       413 ATAGACAAAATGAATTAAATTTACAAAGTGGAAAAAATGAACAAGACATAAATAAAAATG 472

Qy       140 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAA 199
Db       473 AAAAAGGAAAACAGGATATCTCTAATTCCAATGCAGAGAACAAAAAGGATGTAAAAGAAG 532

Qy       200 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 259
Db       533 GAGTGAAAGAATTAGAAGAAAAGAAAAAAGAAGAGAAAATTTCAGATGATCATAAAGTAG 592

Qy       260 AACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAA 319
Db       593 AGGAAAACAAAAAATCTGATGACCATAAAGTTGAGGAAAACAAAAAATCTGATGATCATA 652

Qy       320 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 379
Db       653 AAGTTGAGGAAAACAAAAAATCTGATGATCATAAAATAGAAGAAGTAAAAAAAGTAGAGG 712

Qy       380 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 439
Db       713 AACATGAAGAAGACGAAGAAGAGGATAAGAAAGAAAAAAAATCAGAAAACAAAAATAAAG 772

Qy       440 GACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 499
Db       773 ATGAAAATAAAGATGAAAATGATGAAGATAATGATGAAATAAGTGATGAAGATGAAGTTG 832

Qy       500 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 559
Db       833 ATGATGATGTTGAAGAAGATAAAAATGAAAATGATGATATTGATGATGATAAAAAGGAAA 892

Qy       560 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGC 619
Db       893 CAGATAAAACACATTTGGAAGAAGAGGAAAATGAAATAATTGAAAAGGAATTTAGTGATA 952

Qy       620 AGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGG 679
Db       953 AGAAAAAGAATGGTAAAAATAAAGATACTAAGAAGGAAAAAAGTAAAGATACTGAGAAGG 1012

Qy       680 AACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAAC 739
Db      1013 AAAAAAGTAAAGATATAGAGAAGGAAAAAAGTAAAGATAAAGAGAAGGAAAAAAGTAAAG 1072

Qy       740 TTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACA 799
Db      1073 ATAAAGAGAAGGAAAAAGGTAAAGATAAAGAGAAGGAAAAAAGTAAAGATATTGAGAAGG 1132

Qy       800 ACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAAT 859
Db      1133 AAAAAGAAAAAGATAAAGATATAGAGAAGGAAAAAAGTAAAGATACTGCGAAGGAAAAAG 1192

Qy       860 CTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGA 919
Db      1193 AAAAAGATAAAGATATAGAGAAAGAAAAAAGCAAAGATATGGAAAAATTAAAGAATAAAC 1252

Qy       920 AAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATA 979
Db      1253 AAAATGATGAAAAAAAGAAAGACGATAACGAAAAGAAAAAAAATGATAAACAAGATATAC 1312

Qy       980 CATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATG 1039
Db      1313 ATGATGATAATGATGATGAAAATGATATGGAAGAAATAGAAGAAAATGATGACGAAGAAG 1372

Qy      1040 AAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAG 1099
Db      1373 ATGAGGATGAAGACATGGAAAACAAAAAAAAAAAAAAAAAAAGGAAAAAATGGAAATGAAA 1432

Qy      1100 ACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTT 1159
Db      1433 ATGGAAATGAAAATGGAAGTGAAAATGGAAATGAAAATGGAAATGAAAATGGAAATGAAA 1492

Qy      1160 ATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAA 1219
Db      1493 ATGAAAATAAAAATGAAAGTGAAAATGAAAATGAAAACGAAAATGAAAACGAAAATGGAA 1552

Qy      1220 TAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAA 1279
Db      1553 ACGAAAATGAAAACGAAAAAGAAAACGAAAAAGATAAAAATATTAAAGAGATTGAAAATG 1612

Qy      1280 AAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAA 1339
Db      1613 TAACAAATGCAAACAAGGAAAACTATGAAAAAATAAATAAAAATTCTGAAATAACAATAA 1672

Qy      1340 AATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAA 1379
Db      1673 CAAAATCAAATATAGATATATATAATAATAATAGGAATAA 1712
```

RESULT 16
AAS46735/c
ID AAS46735 standard; DNA; 6292 BP.
XX
AC AAS46735;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #459.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 459; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis

```
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6292 BP; 1736 A; 99 C; 1304 G; 3153 T; 0 U; 0 Other;

  Query Match           10.2%;  Score 151.2;  DB 4;  Length 6292;
  Best Local Similarity   46.2%;  Pred. No. 6.3e-13;
  Matches  619;  Conservative    0;  Mismatches  708;  Indels   13;  Gaps    3;

Qy         40 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG 99
Db       4548 AAATAAAAAAAATAATTAATACCAAAAAAAAAAAAAAAATAACGAACCGAAAAAATCATTA 4489

Qy        100 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 159
Db       4488 CATAAAAACGATAAAAAAAAAATATTAAAAAAAAAAATTAAAAAATAAATAAAAAAAAAA 4429

Qy        160 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 219
Db       4428 AAACAAAATAATAAACAAAAAAAAAAAAAAAATTTCACAAAAAAATAATAAAAAAAAATT 4369

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db       4368 TTTAAATAAATAAAAAAAAATAAAAAAAAATCGTAAAAAAAAAACGAAAATAAACAAAAA 4309

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAAC----AAGA 335
Db       4308 AAAAAAAAAATTTACAAAAATATAATAAAAAAAAAAAAAAAAAAAAAAAAACTATAAATA 4249

Qy        336 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 395
Db       4248 AAAAAATTTTAAATAATAAAAAAAAAAAAAAAAATAACAAATAAAAAAAAAAAAAAAAATA 4189

Qy        396 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAA 455
Db       4188 CAAAAAAAAAAAAAAAAAAAAAAAATAACGAACAAAAAAAAAAAAAAAATTTTATAAAAAAA 4129

Qy        456 GTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 515
Db       4128 CAATAAAAAAAAAAACTTTTAAATAAATAAAAAAAAAAAAAAAAAATAACCAATAAAAAAA 4069

Qy        516 ACAACAAAGCGATTTAGAAC----AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 571
Db       4068 AAATAAAAAAAATAACGAACCGAAAAAAAAAAAAAATTTTACTAAAAAACGATAAACAAAA 4009

Qy        572 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATT 631
Db       4008 AAAAAAATATAAAAAAAAAAAACAATAACTAAAAAATATTTAAATAAACGAAAAAAAAAA 3949

Qy        632 TAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATA 691
Db       3948 AAAAAATAACAAATAAAAAAAAAAATAAAAAATATAACCAAAAAAAAATAAAAAAAAAAAC 3889

Qy        692 TATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAA 751
Db       3888 TAAAAAAAAAACGATAAAAAAAAATAATTTAAAATAAATAAAAAACAAAAAAAAAATAACA 3829

Qy        752 ATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATA 811
Db       3828 AACAAAATAAAAAAAAAAAAAAAACGAACGAAAAAAAAAAAAAAACTTTATAAAAAAATAA 3769

Qy        812 GTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAA 871
Db       3768 CAAAAAAAAATTAAAAAAATAAATAAATTAAAAAAACGTAATAAACAAAAATAAAAAAAA 3709

Qy        872 ATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAA 931
Db       3708 AACTTTACAAAAAAACGACAAAAAAATATTTTTTAAATAAATAAAAAAAAAAAAAAATAA 3649

Qy        932 TAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAG 991
Db       3648 CAAAAAAAAAAAAAAAAAAAACGATAAAAAAACAATTTTTAAATAAACGAAAAAAAAAA 3589

Qy        992 TAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCT- 1050
Db       3588 AAAAAATAACGAACAACAAAAATAAAAAAAAAACTTTAAAAAAAAATAAAAAAAATATTT 3529

Qy       1051 ----GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATT 1106
Db       3528 TTAAAAAAATAAAAAAACAAAAAAAAATAATAAAAACGAAAAACGAAAAAAAAAATAACC 3469

Qy       1107 TAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGA 1166
Db       3468 AAAAAAAAAAAAAAAACACGATAAAAACAAATTTTAAAATAAATAAATAAAAAAAACGAA 3409

Qy       1167 ACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAA 1226
Db       3408 AATAAACGATATAATATAAAAAACTTTACAAAAAAACAATAAAAAAAAAATAATAAAAAA 3349

Qy       1227 ATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAA 1286
Db       3348 AAAATTTTAAAATAAATAAAAAAAAAAAAAAAAATAACGAAAAAAAAAACCAAAAACGATC 3289

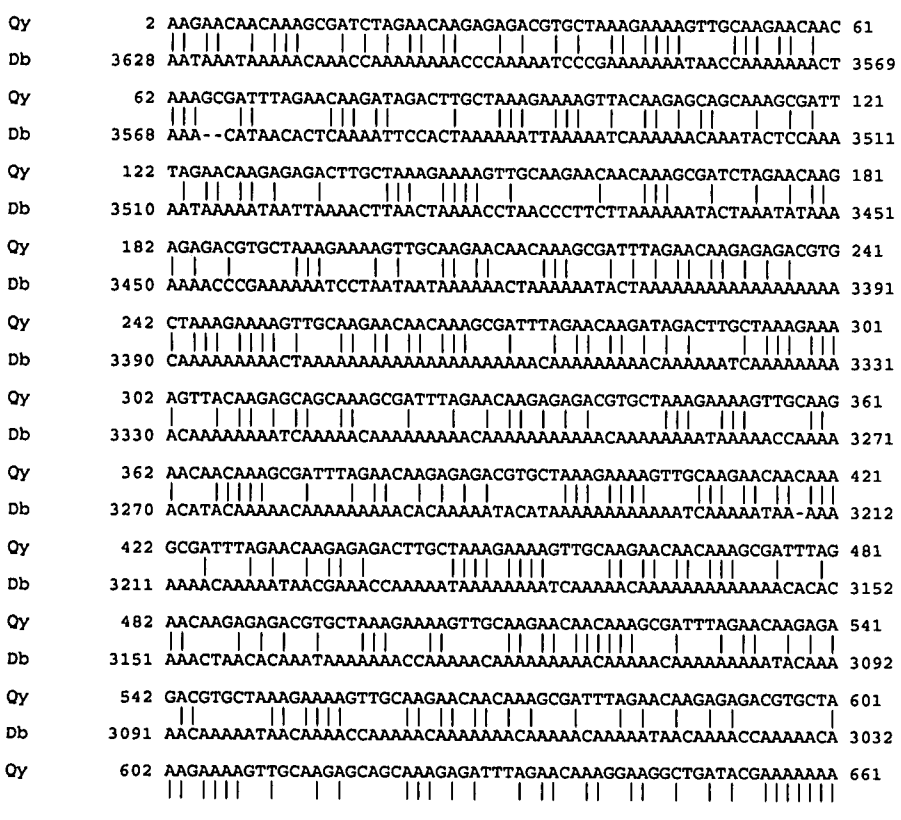
Qy       1287 GGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAA 1346
Db       3288 GAAAAAAAAATTCCTCCTATTACTAATTTTTTAATACTCTTCCATTCCGATCAAAAAAAA 3229

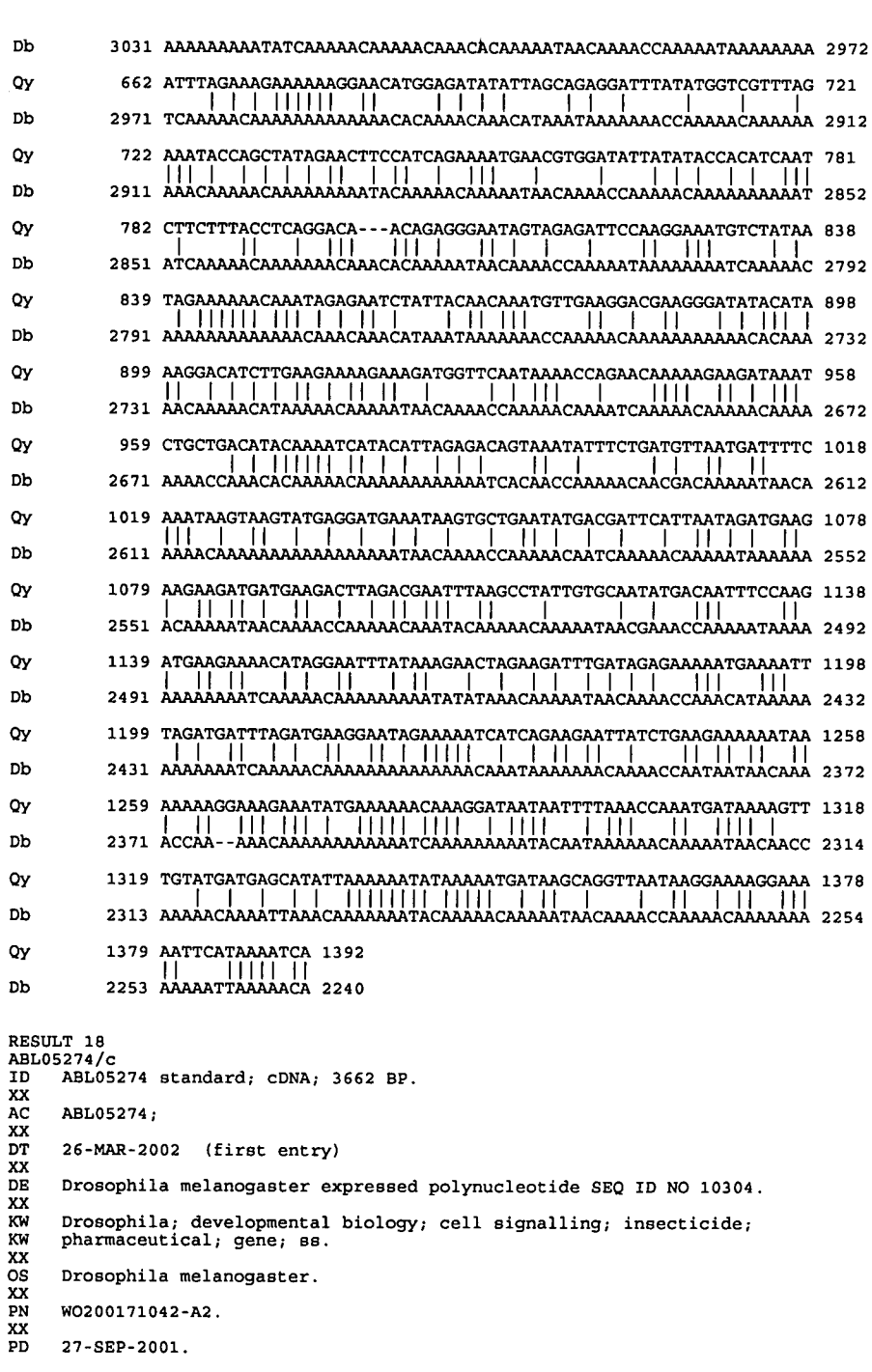
Qy       1347 AAATGATAAGCAGGTTAATA 1366
Db       3228 TAATTAATATAATTTCAAAA 3209
```

```
RESULT 17
AAS46608/c
ID   AAS46608 standard; DNA; 6767 BP.
XX
AC   AAS46608;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #330.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 330; 27pp; English.
XX
```

CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6767 BP; 1446 A; 207 C; 1601 G; 3513 T; 0 U; 0 Other;

Query Match 10.1%; Score 149.2; DB 4; Length 6767;
Best Local Similarity 46.1%; Pred. No. 1.2e-12;
Matches 643; Conservative 0; Mismatches 743; Indels 8; Gaps 4;





RESULT 18
ABL05274/c
ID ABL05274 standard; cDNA; 3662 BP.
XX
AC ABL05274;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10304.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.

```
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61171.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 10304; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3662 BP; 661 A; 941 C; 421 G; 1639 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 144.4;  DB 4;  Length 3662;
  Best Local Similarity   46.7%;  Pred. No. 6e-12;
  Matches  602;  Conservative    0;  Mismatches  676;  Indels   12;  Gaps    4;

Qy         24 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAG 83
Db       1859 AAAAAAAGAAGATGAAATAAACAAGAATCAAGGGAAACCACGAATAATGGAAAAGGAAAG 1800

Qy         84 ACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAA 143
Db       1799 A--AGTAAAGAAGAGGAAAAAAGAAAAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAAA 1742

Qy        144 AGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 203
Db       1741 AGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAA 1682

Qy        204 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 263
Db       1681 GAGGAAAGAAAAAGAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGAAAAA----AAGAAGA 1626

Qy        264 ACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGA 323
Db       1625 GGAAAGAGAAAGAGAAGAGGAAAGAA-----AGAGAGAACATAATAGGAAAAAAGAAGAG 1571

Qy        324 TTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACA 383
Db       1570 GAAAGAAAAAGAGAAGAGAAAAGAAGAAAAGAAGAGGAAAAACGAAAGGAAGAGGAAAGA 1511

Qy        384 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACT 443
Db       1510 AGAAAGGAAGAGGAAAGGAAAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGGAAAGAGGAG 1451

Qy        444 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGA 503
Db       1450 GAAAGAAGAAAGGAAGAGGAAAGAAGAAAAGAGAAAAGAAGAAGAGATGAGAAGAGAAGA 1391

Qy        504 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 563
Db       1390 AGAGAAGAGGAAAAGAGAAAGGAAGAGGAAAGGAAAGAAGAAGAAAGAAGAGAGGAAGCG 1331

Qy        564 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCA 623
Db       1330 GAAAGGAAAGAAGAGGAAAGGAAAGCAGAGGAAAGAAGAAAGAAAGAGGAAAGAAGAAGA 1271

Qy        624 AAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACA 683
Db       1270 GAAGAGAAAAGAAGAAGGGAAGAGA-AAAGAAGAAGGGAAGAAGAGGAAAGAAGGAAAGA 1212

Qy        684 TGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCC 743
Db       1211 AGAGGAAAGAAGAGAAGAGGAGGAAAAAAGAAAAGAAGAGGAAAGAAGGAAAGATGAGGA 1152

Qy        744 ATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAG 803
Db       1151 AAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAG 1092

Qy        804 AGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTAT 863
Db       1091 AGAAGAGGGAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAGAGAAGAGGA 1032

Qy        864 TACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGA 923
Db       1031 AAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAGACGAGGAACGAAGAAGAGA 972

Qy        924 TGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATT 983
Db       971 AGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAGGA 912

Qy        984 AGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAAT 1043
Db       911 ACGAAGAAAAGAAGAGGAGAGAAAAAGAGAAGAGGAAAAAAGAAAAGAAGAGGAGAGAAA 852

Qy       1044 AAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGA 1103
Db        851 AAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAGGAACGAAGAAAAGA 792

Qy       1104 ATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAA 1163
Db        791 AGAGGAGAGAAAAAGAGAAGAGGAAAAAAGAAAAGAAGAGAAGAGAAAAAGAGAAGAGGA 732

Qy       1164 AGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGA 1223
Db        731 AAAAAGAAAAAAAGAGGAGAGAAAAAGAGAAGAGGAAAAAAGAAAAGAAGATGAGAGAAA 672

Qy       1224 AAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAAC 1283
Db        671 AAGAGAAGAGGAAAAAAGAAAAGAAGAGGAAAAAAGAAAGGAGGAGGAAAGAAAAGAAGA 612

Qy       1284 AAAGGATAATAATTTTAAACCAAATGATAA 1313
Db        611 GGAAAGAAAGAAGAAAGAGACAGAGGAAAA 582


RESULT 19
AAS45347/c
ID   AAS45347 standard; DNA; 9539 BP.
XX
AC   AAS45347;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated complementary DNA associated with cell cycle #26.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
```

```
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002945.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analyzing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle.
XX
PS   Claim 1; SEQ ID NO 52; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers
XX
SQ   Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;

  Query Match             9.4%;  Score 139.4;  DB 4;  Length 9539;
  Best Local Similarity   45.6%;  Pred. No. 2.9e-11;
  Matches  607;  Conservative    0;  Mismatches  716;  Indels    8;  Gaps    3;

Qy         72 AGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGA 131
Db       1343 AAAATAAAAAAAAACTAAAAAAACCTTCTATCTATAAACGCAACAAAAATAAAAAAAAAA 1284

Qy        132 GAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGC 191
Db       1283 AAAAATCAAACCCGAAAACCATAAAATTTCTAAAAAAATTAAAAAACAAAATACAACAAA 1224

Qy        192 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAA 251
Db       1223 TAAAAATAAAATTAAAAAACAAAAAATAATAAACCCCAAAATCCTAAATAAAATAACCAA 1164

Qy        252 GTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGA 311
Db       1163 AAAACAAAAAACTCTAATC-CTACATACCATCCCACCCTCACCCAAACTCTTCTAAAAAA 1105

Qy        312 GCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 371
Db       1104 AAAAAAAAAAAAAAACGAAAAAACGAAAAAAATAAAAAAAAATAAAAAAAAAAAAAAAA 1045

Qy        372 CGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 431
Db       1044 AAAATAAAAAAAAAAAAAAAAAAATAAACAAAAAAAAAAAAAAAAAAAATAAAAAAATTAA 985

Qy        432 ACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 491
Db        984 AAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAATAAAAAATAAATATAAACAAAAAAA 925

Qy        492 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 551
Db        924 AAAAACCGAAAATAATAACTCACACCTATAATCCCAATACTTTAAAAAACTAAAATAAAA 865

Qy        552 AGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 611
Db        864 AAATCACTTAAAACCAAAAATTCAAAACCAACCTAAACAACAAAATCATACCCCATTTCT 805

Qy        612 GCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAG 671
Db        804 ACAAAAAAAAAACAAAAACTAACCAAACCTAATAACACACACCTATAATCCCAACTACTC 745

Qy        672 AAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGC 731
Db        744 AAAAAACTAAAACAAAAAAATCACTACCGAAAAACAATA------ATTACAATAAACCGA 691

Qy        732 TATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACC 791
Db        690 AATCGTACCACTACACTCCAACCTAACAACAAAAAAAACCCTACCTCAAAAAAAAATATA 631

Qy        792 TCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAA 851
Db        630 TATAAATAAAAAAAAAAAAAAATAAAATATATAAAAAAAATAAAAAAAAAAAATAAAAACA 571

Qy        852 TAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGA 911
Db        570 AAAAAAAAAAAAAAACAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAATCAAAAAA 511

Qy        912 AGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACA 971
Db        510 AAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAACAAAAAAAAATAAAAAAAAAATAAAAAA 451

Qy        972 AAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTA 1031
Db        450 AAAAAAAATAAAAAAAAAAA-AAACAAAAAAAAAAATAAAAAAAAAAAAAAACGAAAAAAA 392

Qy       1032 TGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGA 1091
Db        391 ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAA 332

Qy       1092 AGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACAT 1151
Db        331 AAACAAATAAAAATAAAAAAAAAAACGAAAAAAAAAAAAAAAAAAAAAAAAACAAACGACAAAA 272

Qy       1152 AGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGA 1211
Db        271 AAAAAAAAACGAAAAAAAAAATAAAAAAAAAAAACACACAAAAATAAAACGAAAAAATACA 212

Qy       1212 TGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAA 1271
Db        211 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAAAAT 152

Qy       1272 ATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCA 1331
Db        151 AAAAAAAAAAAAAAAAATAAAAAAATAAAAAAAAAAAATATAAAAACTAAAATAATAAAAT 92

Qy       1332 TATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATC 1391
Db          91 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAATAAAAAAAAATAAAAAAAAAAATTTC 32

Qy       1392 ATTGTTTCATA 1402
Db          31 TATATCTAAAA 21


RESULT 20
ABK28180/c
ID   ABK28180 standard; DNA; 9539 BP.
XX
AC   ABK28180;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated complementary genomic DNA #27.
```

```
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-090046/12.
XX
PT   New nucleic acids or oligomers, useful for diagnosing or treating
PT   diseases associated with DNA transcription, e.g. immunological disorders,
PT   Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 54; 32pp; English.
XX
CC   The invention relates to a nucleic acid, which comprises a segment of the
CC   chemically pretreated DNA of genes associated with DNA transcription from
CC   one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC   or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC   to the chemically pretreated DNA of genes associated with DNA
CC   transcription. The set of oligomer probes are useful for detecting the
CC   cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC   in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC   diagnosing or treating diseases associated with DNA transcription
CC   (particularly with the methylation status), e.g. adenosine deaminase
CC   deficiency, viral infection, retroviral infection, Sezary syndrome,
CC   haematological disorders, immunological disorders, Werner syndrome,
CC   tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC   neurological disorders, neurodegenerative disorders, Waardenburg
CC   syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC   infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC   disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC   or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC   associated genomic DNA molecules of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification but
CC   was obtained in electronic format directly from the European Patent
CC   Office
XX
SQ   Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;

  Query Match             9.4%;  Score 139.4;  DB 6;  Length 9539;
  Best Local Similarity   45.6%;  Pred. No. 2.9e-11;
  Matches  607;  Conservative    0;  Mismatches  716;  Indels    8;  Gaps    3;

Qy          72 AGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGA 131
               | || || | | |   |  |||     | | | |   ||||    | | | || || |
Db        1343 AAAATAAAAAAAAACTAAAAAAACCTTCTATCTATAAACGCAACAAAAATAAAAAAAAAA 1284

Qy         132 GAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGC 191
                 | | |       |||||     ||        |||     |  | |||||  | | |
```

```
Db        1283 AAAAATCAAACCCGAAAACCATAAAATTTCTAAAAAAATTAAAAAACAAAATACAACAAA 1224

Qy         192 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAA 251
               |||| | ||  |  || ||||| |||           ||     |  |  | |  ||
Db        1223 TAAAAATAAAATTAAAAAACAAAAAATAATAAACCCCAAAATCCTAAATAAAATAACCAA 1164

Qy         252 GTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGA 311
                  ||| ||   | || |  |  | | ||    |  ||  |  ||       || || |
Db        1163 AAAACAAAAAACTCTAATC-CTACATACCATCCCACCCTCACCCAAACTCTTCTAAAAAA 1105

Qy         312 GCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 371
                 |  |||  |   ||| ||  || |   ||| ||||     || || || |||
Db        1104 AAAAAAAAAAAAAAACGAAAAAACGAAAAAAATAAAAAAAAATAAAAAAAAAAAAAAAA 1045

Qy         372 CGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 431
                 | | || || | | |     |||| ||||     || || || | |   | | |
Db        1044 AAAATAAAAAAAAAAAAAAAAAAATAAACAAAAAAAAAAAAAAAAAAAATAAAAAAATTAA 985

Qy         432 ACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 491
               | || | | |       ||  | ||     || || || |||   || || |  || | |
Db         984 AAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAATAAAAAATAAATATAAACAAAAAAAA 925

Qy         492 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 551
               |   |  || | ||      | | | | ||  | | |     || | |    ||
Db         924 AAAAACCGAAAATAATAACTCACACCTATAATCCCAATACTTTAAAAAACTAAAATAAAA 865

Qy         552 AGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 611
               | |  | ||  ||  | ||       |   |    || |   | |  ||    |   |
Db         864 AAATCACTTAAAACCAAAAATTCAAAACCAACCTAAACAACAAAATCATACCCCATTTCT 805

Qy         612 GCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAG 671
               |||  |  | | | | | | |  |    |  | || | ||          |
Db         804 ACAAAAAAAAAACAAAAACTAACCAAACCTAATAACACACACCTATAATCCCAACTACTC 745

Qy         672 AAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGC 731
               ||||||  || |   | | ||      | ||  |  |||       ||| ||  | | |
Db         744 AAAAAACTAAAACAAAAAAATCACTACCGAAAAACAATA------ATTACAATAAACCGA 691

Qy         732 TATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACC 791
               || | ||  | |       ||   |     | |  | |  | || ||||       ||
Db         690 AATCGTACCACTACACTCCAACCTAACAACAAAAAAAACCCTACCTCAAAAAAAAATATA 631

Qy         792 TCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAA 851
               |    | || | |   || | | |  |   || |||       | || | ||| | | |
Db         630 TATAAATAAAAAAAAAAAAAAATAAAATATATAAAAAAAATAAAAAAAAAAAATAAAAACA 571

Qy         852 TAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGA 911
               | | ||   |  |  | |||      ||  || ||    | | | | |||  |     |
Db         570 AAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAATCAAAAAA 511

Qy         912 AGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACA 971
               | |||| ||| |       || ||||  | |||||||| ||   |||       | | | |
Db         510 AAAAAAAAAAAAATAAAAAAAAAAAAAAAAAACAAAAAAAAATAAAAAAAAAATAAAAAAA 451

Qy         972 AAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTA 1031
               |||  | | |  | | | |  ||| |       |  | ||  |     ||| |  ||  |
Db         450 AAAAAAAATAAAAAAAAAAA-AAACAAAAAAAAAAATAAAAAAAAAAAAAAACGAAAAAAAA 392

Qy        1032 TGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGA 1091
                  |  ||| ||      || | |  |   |  | | |  || || || |  || |
Db         391 ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAA 332

Qy        1092 AGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACAT 1151
               | ||  | |  |||  ||     |  |   || |  | ||      || |  ||  | | |
Db         331 AAACAAATAAAAATAAAAAAAAAAACGAAAAAAAAAAAAAAAAAAAAAAAACAAACGACAAAA 272

Qy        1152 AGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGA 1211
               |  ||    | || ||  |  | |  |     | | | |  ||  |    ||  |  || |
Db         271 AAAAAAAAACGAAAAAAAAAATAAAAAAAAAAACACACAAAAATAAAACGAAAAAATACA 212

Qy        1212 TGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAA 1271
                  ||  || | |||||  |  | || ||  |   || ||||| |||||||   || |
Db         211 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAAAAT 152
```

```
Qy       1272 ATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCA 1331
              | |  |||||| |||  | || |||   |||  |||  ||||||  |   || || |
Db        151 AAAAAAAAAAAAAAAAATAAAAAAATAAAAAAAAAAAATATAAAAACTAAAATAATAAAAT 92

Qy       1332 TATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATC 1391
               |  |||||| | |||||  | ||  |    |||||  ||||  |||||   | ||  ||
Db         91 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAATAAAAAAAAATAAAAAAAAAAAATTTC 32

Qy       1392 ATTGTTTCATA 1402
              | | | | |
Db         31 TATATCTAAAA 21
```

RESULT 21
AAX33181
ID AAX33181 standard; DNA; 6644 BP.
XX
AC AAX33181;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRx-ires-bsr.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Cowpox virus.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 38-41; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRx-ires-bsr, which contains the cowpox virus bsr gene, and is used in an
CC example from the present invention
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;

Query Match 9.2%; Score 136.4; DB 2; Length 6644;
Best Local Similarity 49.6%; Pred. No. 7.9e-11;
Matches 350; Conservative 0; Mismatches 356; Indels 0; Gaps 0;

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||| || || |||    |   | || || | | |       ||| ||||     || || ||
Db       3738 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3797

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||   |   | || || | | |       ||| ||||  | || |  |  |||  |
Db       3798 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3857

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              | || || | | |       ||| ||||     || || || |||   |   | || ||
Db       3858 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3917

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              | | |       ||| ||||     || || || |||   |   | ||.|| | | |
Db       3918 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3977

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||  ||||     || || || |||   |   | || || | | |      ||| ||
Db       3978 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4037

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||   | || | | |  |||   |   | || || | | |      ||| ||||     ||
Db       4038 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4097

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              || || |||    |   | || || | | |       ||| ||||     || || || ||
Db       4098 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4157

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              |   |  | || || | | |      ||| ||||     || || || |||   |  |
Db       4158 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4217

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              || || | | |      ||| ||||     || || || |||   |   | || || | |
Db       4218 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4277

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              | |       ||| ||||     || || || |||   |   | || || | | |
Db       4278 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4337

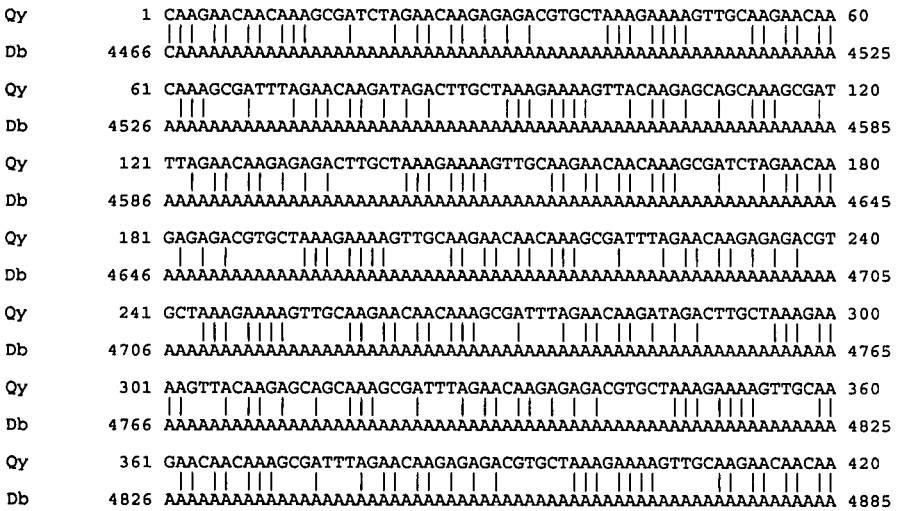
Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||| ||||     || |  |  ||| | |   | || |||  ||      | |  ||||||
Db       4338 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4397

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706
              ||   | |||  ||||||  || |   | |   |||  |  || | ||
Db       4398 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAATTGGTCGAAGCTT 4443
```

RESULT 22
AAX33182
ID AAX33182 standard; DNA; 7372 BP.
XX
AC AAX33182;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRx-Bcl-xl-bsr.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9913073-A2.

XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 2; Page 41-45; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and is used in an
CC example from the present invention
XX
SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

Query Match 9.2%; Score 136.4; DB 2; Length 7372;
Best Local Similarity 49.6%; Pred. No. 7.9e-11;
Matches 350; Conservative 0; Mismatches 356; Indels 0; Gaps 0;





RESULT 23
AAX33180
ID AAX33180 standard; DNA; 7797 BP.
XX
AC AAX33180;
XX
DT 25-JUN-1999 (first entry)
XX
DE Cowpox virus bsr full length gene sequence.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Cowpox virus.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 34-38; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an

```
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the cowpox virus bsr gene which
CC   is used in an example from the present invention
XX
SQ   Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other;

  Query Match             9.2%;  Score 136.4;  DB 2;  Length 7797;
  Best Local Similarity   49.6%;  Pred. No. 7.8e-11;
  Matches  350;  Conservative    0;  Mismatches  356;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||| || || |||   |   | || || | | |       ||| ||||     || || ||
Db       4891 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4950

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||   |   | || || | | |       ||| ||||   | || |  |  |||   |
Db       4951 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5010

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              | || || | | |       ||| ||||     || || || |||   |   | || ||
Db       5011 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5070

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              | | |      ||| ||||     || || || |||   |   | || || | | |
Db       5071 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5130

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||  ||||     || || || |||   |   | || || | | |       ||| ||
Db       5131 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5190

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||   | || |  |  |||   |   | || || | | |       ||| ||||     ||
Db       5191 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5250

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              || || |||   |   | || || | | |       ||| ||||     || || || ||
Db       5251 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5310

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              |   |   | || || | | |       ||| ||||     || || || |||   |   |
Db       5311 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5370

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              || || | |      ||| ||||     || || || |||   |   | || || |
Db       5371 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5430

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              | |      ||| ||||     || || || |||   |   | || || | | |
Db       5431 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5490

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||  ||||     || |  |  ||| | |   | || |||  ||     | |  ||||||
Db       5491 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5550

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706
              ||   | |||  ||||||  || |   | |   |||  |  || | ||
Db       5551 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAATTGGTCGAAGCTT 5596


RESULT 24
AAX33184
ID   AAX33184 standard; DNA; 7996 BP.
XX
AC   AAX33184;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-Bcl 2-i-hCD 25.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW   bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP004010.
XX
PR   08-SEP-1997;   97JP-00259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell.
XX
PS   Example 3; Page 46-49; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the plasmid
CC   pRx-Bcl 2-i-hCD 25, which contains the human Bcl-2 gene, and is used in
CC   an example from the present invention
XX
SQ   Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 U; 0 Other;

  Query Match             9.2%;  Score 136.4;  DB 2;  Length 7996;
  Best Local Similarity   49.6%;  Pred. No. 7.8e-11;
  Matches  350;  Conservative    0;  Mismatches  356;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||| || || |||   |   | || || | | |       ||| ||||     || || ||
Db       5090 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5149

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||   |   | || || | | |       ||| ||||   | || |  |  |||   |
Db       5150 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5209

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              | || || | | |       ||| ||||     || || || |||   |   | || ||
Db       5210 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5269

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              | | |      ||| ||||     || || || |||   |   | || || | | |
Db       5270 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5329

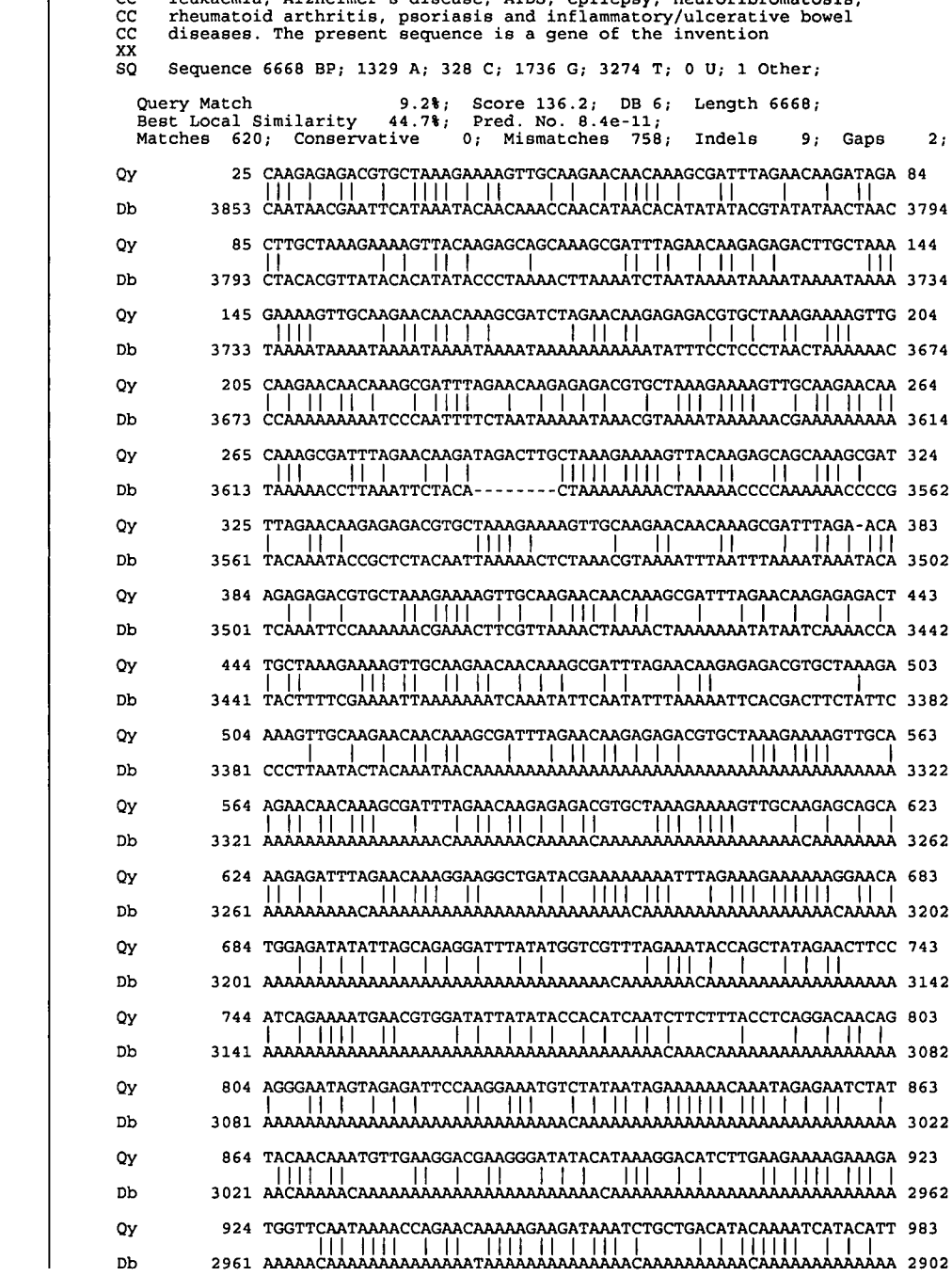
Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||  ||||     || || || |||   |   | || || | | |       ||| ||
Db       5330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5389

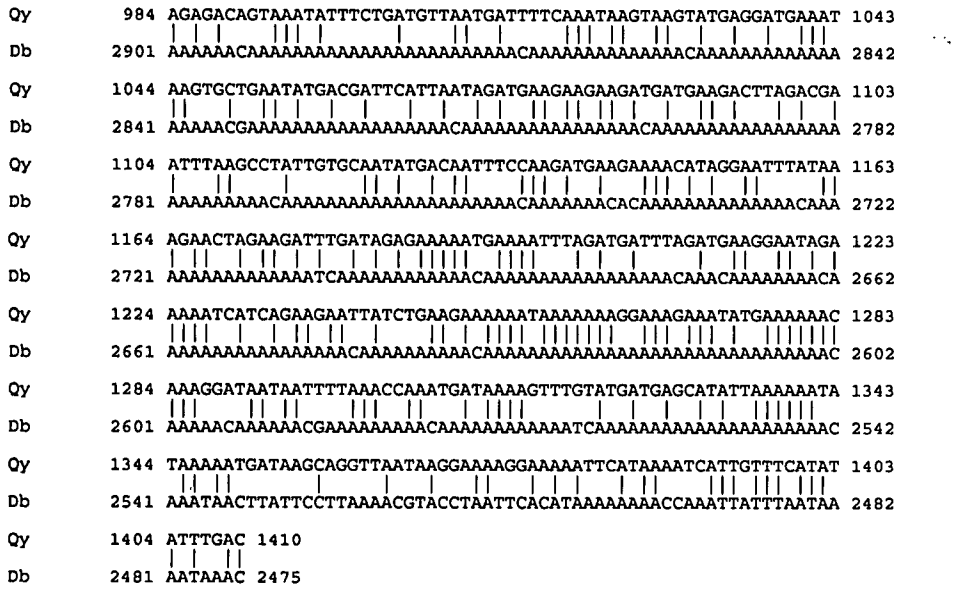
Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||   | || |  |  |||   |   | || || | | |       ||| ||||     ||
Db       5390 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5449
```



RESULT 25
ABL33697/c
ID ABL33697 standard; DNA; 6668 BP.
XX
AC ABL33697;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1670.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;

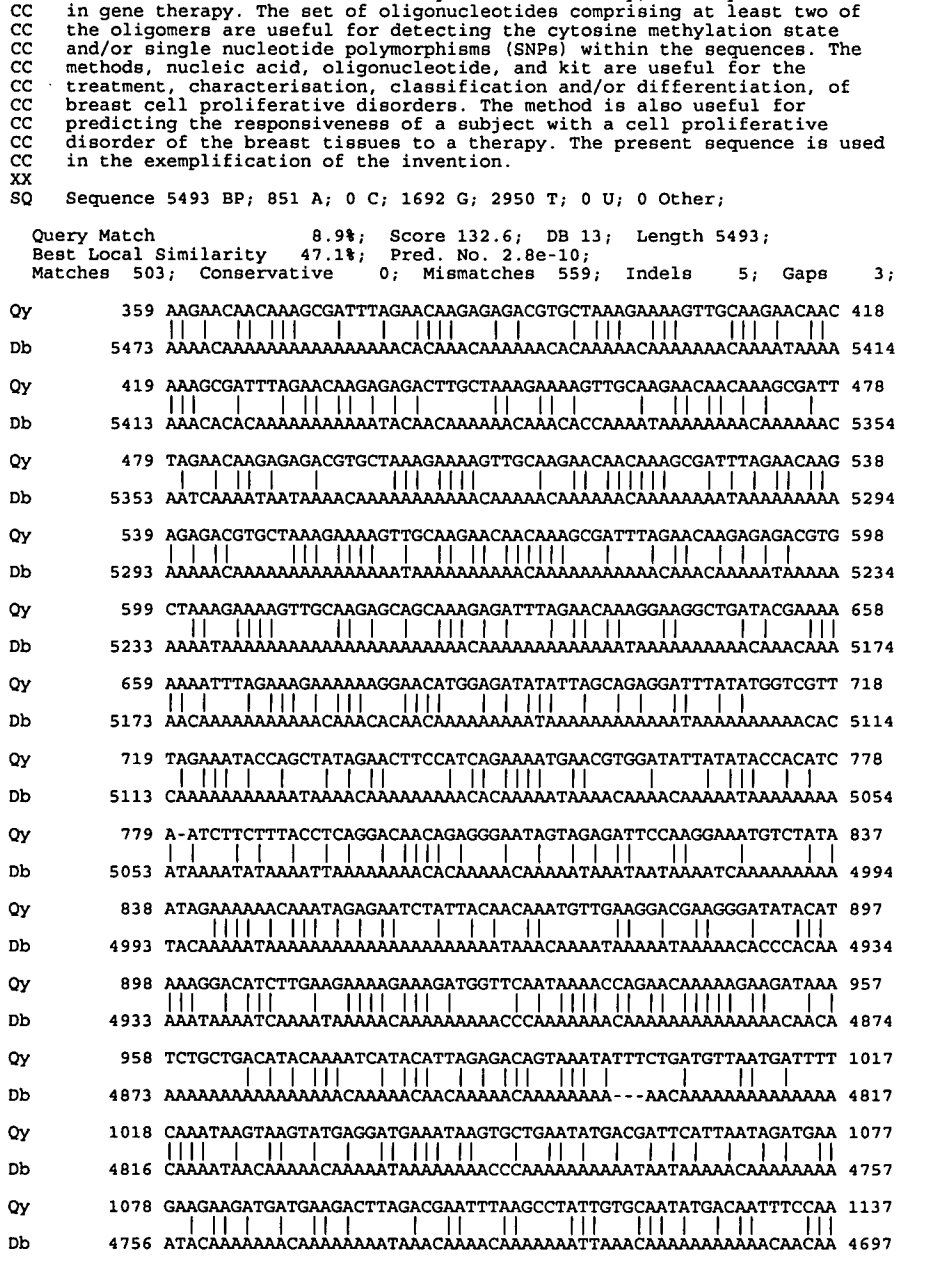
Query Match 9.2%; Score 136.2; DB 6; Length 6668;
Best Local Similarity 44.7%; Pred. No. 8.4e-11;
Matches 620; Conservative 0; Mismatches 758; Indels 9; Gaps 2;





RESULT 26
ADS89675/c
ID ADS89675 standard; DNA; 5493 BP.
XX
AC ADS89675;
XX
DT 18-NOV-2004 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:691.
XX
KW ss; cell proliferative disorder; breast; methylation; cytostatic;
KW gene therapy; single nucleotide polymorphism; SNP.
XX
OS Unidentified.
XX
PN WO2004035803-A2.
XX
PD 29-APR-2004.
XX
PF 01-OCT-2003; 2003WO-EP010881.
XX
PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
PI Nimmrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX
PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
PS Claim 25; SEQ ID NO 691; 104pp; English.
XX
CC The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytostatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
SQ Sequence 5493 BP; 851 A; 0 C; 1692 G; 2950 T; 0 U; 0 Other;

Query Match 8.9%; Score 132.6; DB 13; Length 5493;
Best Local Similarity 47.1%; Pred. No. 2.8e-10;
Matches 503; Conservative 0; Mismatches 559; Indels 5; Gaps 3;



```
Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              |  || |||| | | | ||       | ||| |  | |    | | | | | |  |||
Db       4696 AACAAAAAAAAAAAAACAAAAACCCCAAAACCACCAAAAAACACACACACATAAAAAATA 4637

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGA-AAAAAT 1256
                | |  | ||| |   |  || |  |||   | || || ||  |    || | ||||||
Db       4636 AAAAAACAATTAAAATTAAAAAAAACAAACAAAACAAAACAAAAACAAAAAAACAAAAAT 4577

Qy       1257 AAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAG 1316
              || |||| ||| || || |||| |  |  | || |||    ||   ||  ||||||
Db       4576 AACAAAACAAAACCAAAATCAAAACCCCCACCAAAAAAATACCCAATACAAACATAAAAA 4517

Qy       1317 TTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGA 1376
              |       |  |    | ||| ||| | ||| |  |  | |    || ||  |||| |
Db       4516 TAAAACAAAAAAAACAAATAACAAAAAAAAACAAAAATAAAAAAAAAAAAAAAAAAAAAA 4457

Qy       1377 AAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAA 1423
              |||| | | ||||  |     ||| |||| | |   | | ||  |||
Db       4456 AAAAATAAAAAAAAAAAAAAAATCAAATATATAAAAAAAAAAAAAAA 4410


RESULT 27
ABQ15588/c
ID   ABQ15588 standard; DNA; 778 BP.
XX
AC   ABQ15588;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 2179.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP010074.
XX
PR   01-SEP-2000; 2000DE-01043826.
PR   05-SEP-2000; 2000DE-01044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful for
PT   diagnosis and prognosis, comprises selective hybridization of amplicons
PT   from chemically treated DNA.
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one member,
CC   of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC   degree of hybridisation to both classes is determined from the label on
CC   the amplicon. From the ratio of labels hybridised to the two classes of
CC   oligomers, the degree of methylation is calculated. The method is used:
CC   (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC   and of a wide range of diseases, e.g. cancer, disorders of the central
CC   nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC   particularly by detecting mutations or single nucleotide polymorphisms
CC   (SNP's); and (ii) for differentiation of cell or tissue types and for
CC   investigating cell differentiation. The method allows the methylation
CC   status of many C residues to be determined simultaneously. ABQ13410-
CC   ABQ54121 represent genomic DNA sequences used to illustrate the method
CC   for determining the degree of cytosine methylation described in the
CC   disclosure of the invention
XX
SQ   Sequence 778 BP; 111 A; 24 C; 29 G; 614 T; 0 U; 0 Other;

  Query Match             8.9%;  Score 131.2;  DB 6;  Length 778;
  Best Local Similarity   48.4%;  Pred. No. 5e-10;
  Matches  364;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||   | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db        773 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 714

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
                | |   | |       ||  |||| |  |   | | |      |  | |||| |  |
Db        713 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 654

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
                | |  |  ||| | |  ||  |  | ||    |  | |  | |   || |  | |
Db        653 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAATTATAAAAAAAAAAAAAA 594

Qy        818 ATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  |||     || ||| |||||| ||| |    |   |  | || ||||   |
Db        593 ATTACGAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACG 534

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | ||| | |  ||  | |||| ||| |   ||    ||||
Db        533 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAA 474

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
                | ||   |  || | |||     | ||  | ||||  | |       | | |  ||| |
Db        473 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 414

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |    ||  ||| | ||| || ||| || ||  |  ||| ||       | |
Db        413 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 354

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   | || || || || || |  | |   | |    ||     |
Db        353 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTACGAAAAAAAAAT 294

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||           || ||||  || |||   | | |  || | || |
Db        293 TACGAAAAAAAAAAAAAAATTATAAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAA 234

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |   | |||  ||||   | || |   | |  ||  || | |  |  |  | | ||
Db        233 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAA 174

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | ||| | |||  | || ||||
Db        173 AAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAAATT 114

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
                ||||  |||  | |  |      |  |  |  ||  |||||| | |||||     |
Db        113 ATAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTATAAA 54

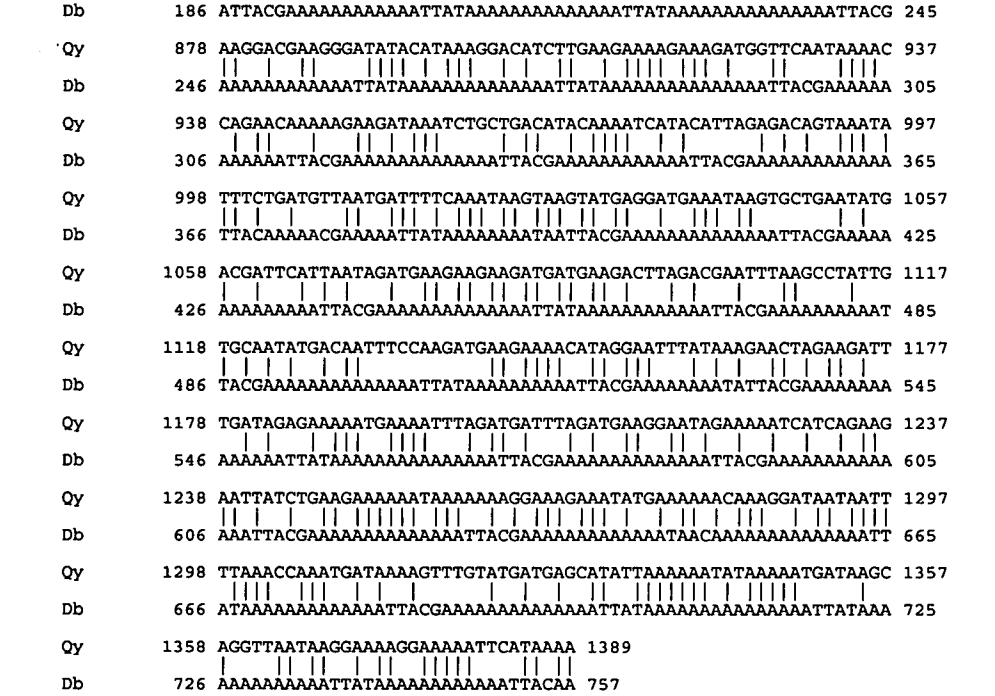
Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db         53 AAAAAAAAAATTATAAAAAAAAAAAAATTACAA 22


RESULT 28
ABQ15589
ID   ABQ15589 standard; DNA; 778 BP.
XX
AC   ABQ15589;
XX
```

DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 2180.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 778 BP; 614 A; 29 C; 24 G; 111 T; 0 U; 0 Other;

Query Match 8.9%; Score 131.2; DB 6; Length 778;
Best Local Similarity 48.4%; Pred. No. 5e-10;
Matches 364; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

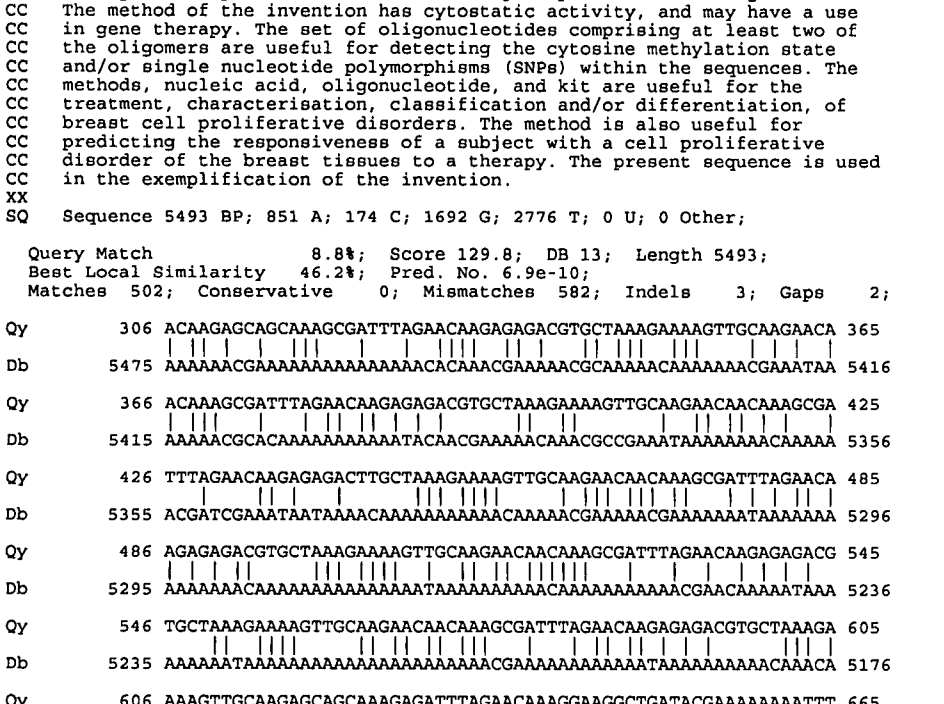




RESULT 29
ADS89401/c
ID ADS89401 standard; DNA; 5493 BP.
XX
AC ADS89401;
XX
DT 18-NOV-2004 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:417.
XX
KW ss; cell proliferative disorder; breast; methylation; cytostatic;
KW gene therapy; single nucleotide polymorphism; SNP.
XX
OS Unidentified.
XX
PN WO2004035803-A2.
XX
PD 29-APR-2004.
XX
PF 01-OCT-2003; 2003WO-EP010881.
XX
PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
PI Nimmrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX
PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell

PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
PS Claim 25; SEQ ID NO 417; 104pp; English.
XX
CC The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytostatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
SQ Sequence 5493 BP; 851 A; 174 C; 1692 G; 2776 T; 0 U; 0 Other;

Query Match 8.8%; Score 129.8; DB 13; Length 5493;
Best Local Similarity 46.2%; Pred. No. 6.9e-10;
Matches 502; Conservative 0; Mismatches 582; Indels 3; Gaps 2;





RESULT 30
AAS46686/c
ID AAS46686 standard; DNA; 7442 BP.
XX
AC AAS46686;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #409.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 409; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7442 BP; 1655 A; 214 C; 1934 G; 3636 T; 0 U; 3 Other;

Query Match 8.8%; Score 129.8; DB 4; Length 7442;
Best Local Similarity 46.7%; Pred. No. 6.8e-10;
Matches 564; Conservative 0; Mismatches 627; Indels 18; Gaps 4;

```
Qy        194 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 253
              || ||   |   |  | || |||   ||  |  || |      |    | ||| ||||
Db       3079 AAAAACCCTCAAACACAAAAAAAAAATATAAAACACTATCCCAATACACAAAAAAAAAAA 3020

Qy        254 TGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGC 313
                ||   || ||| |||     || |  | |  | |  ||| ||||   | || |
Db       3019 CTAAAATTTAAAAAAACGAAACTTAAAATAAAAACCCTCAAAAATAAAATAAAAAAAACA 2960

Qy        314 AGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 373
              |||| | ||       |  | |  |      ||  ||       ||   ||||  | |
Db       2959 CCCAAACCTATCCCACTCCTAACATCCCAAAACAACTAATACAATAATTCCAACCTAACA 2900

Qy        374 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 433
              |      | || | | |     |||  ||       || |     |||       | ||
Db       2899 AAACTATAAAAAAAAAAAAATATAATAAATCCGATAAAAACTTCGAAAAACTAAAAAAAA 2840

Qy        434 AAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 493
                 |||    ||| |||   |   |   |   |||   |  | || |  | | |
Db       2839 CGATACTACTAAAAAAAAAAATACTAATAATTCTCTAAAATTAAAAAAAAAATAAAAAAA 2780

Qy        494 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAG 553
              |||| ||||  |  || || || |||   || || || || |   |     | ||
Db       2779 TCAAAAAAAAAAAAAATCTAAAAAAAAAAAAATAAATCTAAAAAAAAAATAAATAATTTAAT 2720

Qy        554 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 613
              ||||     || || ||       |   | ||  | |       |   ||| | ||    |
Db       2719 AAAAAAAAAAAAAAAAAATCTAAAAAAAAAAAAACAAAATCTATTAAAAAAAATAAAAATC 2660

Qy        614 AAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAA 673
               |    |  ||| | |  |  || ||   ||      ||   ||||||||      |||
Db       2659 TAATAAAAAAAATAAAAATCTAAAAATAAAAACGAAATTTAAAAAAAAAAAAATAAAAATC 2600

Qy        674 AAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTA 733
              |||  || |   | | || | |  | |  | | | |   |     | |||    |  |
Db       2599 TAAAAAAAAAAACAAAAATCTAAAAAAAAAAAATAAAAATCTAAAAAAAAAAACGAAAATT 2540

Qy        734 TAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTC 793
              || ||      |  | |||   ||      | | || | | |  |  |       |  |  ||
Db       2539 TAAAAAAAAAAATAAAAATCTAAAAAAAAAAAATAAAAATCTAAAAATAAAAATAAAATTC 2480

Qy        794 AGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATA 853
               | || | |   || |     | |    ||  |||  |||| || | ||||   ||||
Db       2479 GAAAAAAAAAAATAAAAAACTAAAAAAAAAAAATAAAAATCTAAAAAAAAAAAACGAAAATC 2420

Qy        854 GAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAG 913
              | ||   || | ||   |     ||  |  ||   || | | |||    | | |  |
Db       2419 TAAAAAAAATAAAAATCTAAAAAAAAAAAATAAAAATCTAAAAAAAAAAAATAAACATATAA 2360
```

```
Qy        914 AAAAGAA----------AGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGC 962
              |||| ||          | |       |||||||    || ||||| || | ||||||
Db       2359 AAAAAAAAATAAAAATTTAAAAAAAAAAATAAAAATCTAAAAAAAAAAATAAAAATCTAA 2300

Qy        963 TGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAAT 1022
                | | | | ||   |  |  | | | | |||| || |   |    |   | |  |||
Db       2299 AAATAAAAATAAAATTTAAAAAAAAAAAATAAAAATCTAAAAAAAAATAAAACTCTAAAA 2240

Qy        1023 AAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGA 1082
               ||  || ||  |   | ||| ||    | || ||   |   |  |||| |     || |
Db       2239 AAAAAAATAAAAATCTAAAAAAAAAAAATAAAAATCTAAAAAAAAAAAATAAACATCTAAAA 2180

Qy        1083 AGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGA 1142
               | |  |  || | ||  |  || | ||    | | |  || |  | |||     |    |
Db       2179 AAAAAAATAAAAATTTAAAAAAATAAATAAAAATTTAAAAAAAAA-AATAAAAATCTAAA 2121

Qy        1143 AGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTT--A 1200
               | |||| |||  ||| || ||| ||  | || | |  | | | |||||| ||||| |  |
Db       2120 AAAAAAAATAAAAATCTAAAAAAAAAAAATAAAAATCTAAATAAAAAAATAAAAATCTAAA 2061

Qy        1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
               |  |   | | |   || | ||||| ||  |    ||  |    || ||| | ||||
Db       2060 AAAAAAAAAAAATATCTAAAAAAAAAAACACAATCTAAAAAAATCTAAAAAACACTAAAT 2001

Qy        1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
               |||  || | ||     |||||| |||   || ||  |||||  |||  | ||||  |
Db       2000 AAAAAAACGCAACTCTAAAAAAAAAA----AAAAAAATTAAAAAAAAAAAAAAAATCTAA 1945

Qy        1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
               |  |  |   |  |||||| |||||| | || ||  |  | ||  ||||  |||||
Db       1944 AAAAAAAAAATTTAAAAAAAATCTAAAAATAAAAACCAAATCTAAAAAAAAAAAAAAAA 1885

Qy        1381 TTCATAAAA 1389
               ||| ||||
Db       1884 CTCAAAAAA 1876
```

Search completed: December 20, 2005, 05:54:35
Job time : 1149.77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:28:16 ; Search time 190 Seconds
(without alignments)
247.440 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	117	2	AAR26937	Aar26937 P.falcipa
2	554	100.0	316	2	AAR26941	Aar26941 P.falcipa
3	554	100.0	493	2	AAR26944	Aar26944 P.falcipa
4	553	99.8	456	8	ADO21942	Ado21942 LSA-NRC(H
5	553	99.8	457	8	ADO21920	Ado21920 LSA-NRC(H
6	172	31.0	34	8	ADO21926	Ado21926 Malaria p
7	103	18.6	20	8	ADO21925	Ado21925 Malaria p
8	94	17.0	452	9	ADZ40617	Adz40617 PfCTL/HTL
9	94	17.0	456	7	ADA49437	Ada49437 Multi-epi
10	94	17.0	456	8	ADO24115	Ado24115 Epigene c
11	85	15.3	15	2	AAW85415	Aaw85415 Helper T-
12	85	15.3	15	2	AAW85172	Aaw85172 Helper T-
13	85	15.3	15	7	ADW36112	Adw36112 HLA bindi
14	85	15.3	15	7	ADW35943	Adw35943 HLA bindi
15	82	14.8	829	9	AEA26643	Aea26643 Stress to
16	81.5	14.7	231	4	AAG98701	Aag98701 Human cel
17	81.5	14.7	232	4	AAG98700	Aag98700 Human cel
18	81.5	14.7	238	4	AAG98699	Aag98699 Human cel
19	81.5	14.7	257	4	AAG98698	Aag98698 Human cel
20	81.5	14.7	272	4	AAG98697	Aag98697 Human cel
21	81.5	14.7	379	5	ABP26206	Abp26206 Streptoco
22	81.5	14.7	412	8	ADU69665	Adu69665 S agalact
23	81.5	14.7	412	8	ADV89658	Adv89658 Streptoco
24	81.5	14.7	412	8	ADV83060	Adv83060 Streptoco
25	81.5	14.7	412	8	ADV80911	Adv80911 Streptoco
26	81.5	14.7	485	3	AAY91310	Aay91310 Group B S
27	81.5	14.7	2447	4	ABG22380	Abg22380 Novel hum
28	81	14.6	294	8	ADX79752	Adx79752 Plant ful
29	80.5	14.5	371	9	ADZ72255	Adz72255 Plasmodiu
30	80.5	14.5	903	6	ABU24404	Abu24404 Protein e
31	80	14.4	575	8	ADT56884	Adt56884 Plant pol
32	79.5	14.4	271	5	ABB04736	Abb04736 Human cyl
33	79.5	14.4	282	3	AAY58635	Aay58635 Protein r
34	79.5	14.4	321	4	AAM14001	Aam14001 Peptide #
35	79.5	14.4	321	4	ABB32946	Abb32946 Peptide #
36	79.5	14.4	321	4	AAM26407	Aam26407 Peptide #
37	79.5	14.4	321	4	ABB27775	Abb27775 Human pep
38	79.5	14.4	321	4	ABB18420	Abb18420 Protein #
39	79.5	14.4	321	4	AAM66130	Aam66130 Human bon
40	79.5	14.4	321	4	AAM53747	Aam53747 Human bra
41	79.5	14.4	321	4	ABG47798	Abg47798 Human liv
42	79.5	14.4	321	4	AAM01742	Aam01742 Peptide #
43	79.5	14.4	321	5	ABG35780	Abg35780 Human pep
44	79.5	14.4	361	4	AAM14436	Aam14436 Peptide #
45	79.5	14.4	361	4	AAM13889	Aam13889 Peptide #

ALIGNMENTS

RESULT 1
AAR26937
ID AAR26937 standard; peptide; 117 AA.
XX
AC AAR26937;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum liver-stage specific antigen non-repeat region.
XX
KW Malaria; LSA-NR; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW paludism.
XX
OS Plasmodium falciparum.
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.
XX
PR 05-FEB-1991; 91FR-00001286.
XX
PA (INSP) INST PASTEUR.
XX
PI Guerinmarchand C, Druilhe P;
XX
DR WPI; 1992-299985/36.
XX
PT Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT vaccination against, treatment of and diagnosis of malaria.
XX
PS Claim 4; Page 55; 81pp; French.
XX
CC A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC gt11 was used to transform E.coli. The expression library was screened
CC with human antisera against antigens of all stages of P. falciparum. The
CC library was rescreened with antibodies affinity- purified on a clone
CC which was able to recognise antibodies specific to the hepatic phase.
CC About 40 clones were detected which produced a characteristic LSA
CC epitope. The clone with the largest insert encoded LSA-R-NR containing a
CC 12-repeat region (LSA-R) followed by a non-repeat region (LSA-NR). The
CC sequence given here is a preferred polypeptide of the invention carrying
CC a T cell epitope typical of liver-stage P.falciparum. See AAR26919-
CC R26935, AAR26937 and AAR26939. (Updated on 25-MAR-2003 to correct PN
CC field.)

```
XX
SQ   Sequence 117 AA;

  Query Match             100.0%;  Score 554;  DB 2;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-55;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 70

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         71 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 117


RESULT 2
AAR26941
ID   AAR26941 standard; protein; 316 AA.
XX
AC   AAR26941;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA-R-NR protein.
XX
KW   Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism; liver stage-specific antigen.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .209
FT                   /label= repeat_region
FT                   /note= "contains 12 x 17mer repeats"
FT   Region          210. .316
FT                   /label= non-repeat_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28115.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Disclosure; Fig 1; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert (950 bases) encoded LSA-R-NR
CC   containing a 12-repeat region followed by a non-repeat region. Preferred
CC   antigenic polypeptides of the invention are derived from the amino acid
CC   sequence of LSA-R-NR. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 316 AA;

  Query Match             100.0%;  Score 554;  DB 2;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-54;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 269

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        270 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 316


RESULT 3
AAR26944
ID   AAR26944 standard; protein; 493 AA.
XX
AC   AAR26944;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA gene C-terminal region.
XX
KW   Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW   paludism; liver stage-specific antigen.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          13. .213
FT                   /label= repetitive_region
FT   Region          214. .493
FT                   /label= non-repetitive_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28119.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 2; Fig 8-10; 81pp; French.
XX
CC   The 3' part of the P.falciparum liver-stage specific antigen (LSA) gene
CC   codes for a polypeptide sequence which carries a T cell epitope
CC   characteristic of a protein produced in hepatocytes infected with
CC   P.falciparum. The polypeptide can be used in the preparation of vaccines
CC   against malaria. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 493 AA;

  Query Match             100.0%;  Score 554;  DB 2;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-54;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 273

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        274 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 320
```

```
RESULT 4
ADO21942
ID   ADO21942 standard; protein; 456 AA.
XX
AC   ADO21942;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H) construct protein derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H) construct.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   N-PSDB; ADO21941.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 4; SEQ ID NO 26; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to a C-terminal His6 tag.
XX
SQ   Sequence 456 AA;

  Query Match             99.8%;  Score 553;  DB 8;  Length 456;
  Best Local Similarity   99.1%;  Pred. No. 7.4e-54;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        165 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 224

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 271


RESULT 5
ADO21920
ID   ADO21920 standard; protein; 457 AA.
XX
AC   ADO21920;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H)Mut construct protein derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H)Mut construct; mutant; mutein.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 433
FT                   /note= "Residue inserted relative to wild-type P.
FT                   falciparum T5 epitope"
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   N-PSDB; ADO21919.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 5; SEQ ID NO 4; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to an amino acid insertion
CC   within the T5 P. falciparum epitope and a C-terminal His6 tag.
XX
SQ   Sequence 457 AA;

  Query Match             99.8%;  Score 553;  DB 8;  Length 457;
  Best Local Similarity   99.1%;  Pred. No. 7.5e-54;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        165 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 224

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 271


RESULT 6
ADO21926
ID   ADO21926 standard; peptide; 34 AA.
XX
AC   ADO21926;
XX
DT   12-AUG-2004  (first entry)
XX
```

```
DE   Malaria parasite P. falciparum LSA-1 LSA-ter epitope peptide.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; epitope; LSA-ter.
XX
OS   Plasmodium falciparum.
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 2; SEQ ID NO 10; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the malaria parasite P. falciparum LSA-1 LSA-ter epitope
CC   peptide of the invention.
XX
SQ   Sequence 34 AA;

  Query Match             31.0%;  Score 172;  DB 8;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 5e-12;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHKGHL 90
              ||||||||||||||||||||||||||||||||||
Db          1 NSRDSKEISIIEKTNRESITTNVEGRRDIHKGHL 34


RESULT 7
ADO21925
ID   ADO21925 standard; peptide; 20 AA.
XX
AC   ADO21925;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Malaria parasite P. falciparum liver stage antigen-1 NR epitope peptide.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; epitope; NR.
XX
OS   Plasmodium falciparum.
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 2; SEQ ID NO 9; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the malaria parasite P. falciparum LSA-1 NR epitope peptide of
CC   the invention.
XX
SQ   Sequence 20 AA;

  Query Match             18.6%;  Score 103;  DB 8;  Length 20;
  Best Local Similarity   95.0%;  Pred. No. 0.00017;
  Matches   19;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 TKKNLERKKEHGDILAEDLY 24
              ||||||||||||||:||||||
Db          1 TKKNLERKKEHGDVLAEDLY 20


RESULT 8
ADZ40617
ID   ADZ40617 standard; protein; 452 AA.
XX
AC   ADZ40617;
XX
DT   30-JUN-2005  (first entry)
XX
DE   PfCTL/HTL/N multi-epitope construct protein.
XX
KW   multi-epitope; immune stimulation; HIV infection; anti-HIV;
KW   immunostimulant; vaccine; immunogenicity.
XX
OS   Synthetic.
XX
PN   WO2005033265-A2.
XX
PD   14-APR-2005.
XX
PF   26-APR-2004; 2004WO-US012732.
XX
PR   25-APR-2003; 2003US-0465229P.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Chesnut RW,  Newman MJ,  Livingston BD;
XX
DR   WPI; 2005-285416/29.
DR   N-PSDB; ADZ40618.
XX
PT   New multi-epitope polynucleotide, useful for designing vaccines that
PT   provide increased immunogenicity, and for inducing immune responses
PT   against HIV in an individual.
XX
PS   Disclosure; Fig 18I; 261pp; English.
XX
CC   The invention relates to a novel multi-epitope polynucleotide. The
CC   invention further comprises: a composition comprising the polynucleotide;
CC   a cell comprising the polynucleotide; a method for inducing an immune
CC   response against human immunodeficiency virus (HIV) in an individual,
```

CC comprising administering the polynucleotide, composition or cell to the
CC individual; and a method for making the polynucleotide, composition or
CC cell. The multi-epitope polynucleotide has anti-HIV and immunostimulant
CC activities. The polynucleotide is useful for designing vaccines that
CC provide increased immunogenicity, and for inducing an immune response
CC against HIV. This sequence represents an PfCTL/HTL/N multi-epitope
CC construct protein of the invention.
XX
SQ Sequence 452 AA;

Query Match 17.0%; Score 94; DB 9; Length 452;
Best Local Similarity 85.7%; Pred. No. 0.12;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 30 PAIELPSENERGYYIPHQSSL 50
| ||||||||||||||||||
Db 119 PGPGLPSENERGYYIPHQSSL 139

RESULT 9
ADA49437
ID ADA49437 standard; protein; 456 AA.
XX
AC ADA49437;
XX
DT 20-NOV-2003 (first entry)
XX
DE Multi-epitope construct #21.
XX
KW multi-epitope; immunogenic; epitope; major histocompatibility complex;
KW MHC class I; MHC class II; junctional epitope.
XX
OS Synthetic.
OS Plasmodium falciparum.
XX
PN US2002119127-A1.
XX
PD 29-AUG-2002.
XX
PF 27-JUN-2001; 2001US-00894018.
XX
PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.
XX
PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.
XX
PI Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;
XX
DR WPI; 2003-615704/58.
DR N-PSDB; ADA49438.
XX
PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.
XX
PS Disclosure; Fig 18I; 78pp; English.
XX
CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer
CC residues between two CEN, and selecting the constructs having less
CC junctional epitopes. The method is useful for designing a multi-epitope
CC construct having multiple epitope nucleic acid. The method avoids or
CC minimises the occurrence of junctional epitopes and maximises the
CC immunogenicity and/or antigenicity of multi-epitope vaccines. The present
CC sequence represents the amino acid sequence of a multi-epitope construct.
XX
SQ Sequence 456 AA;

Query Match 17.0%; Score 94; DB 7; Length 456;
Best Local Similarity 85.7%; Pred. No. 0.12;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 30 PAIELPSENERGYYIPHQSSL 50
| ||||||||||||||||||
Db 119 PGPGLPSENERGYYIPHQSSL 139

RESULT 10
ADO24115
ID ADO24115 standard; protein; 456 AA.
XX
AC ADO24115;
XX
DT 01-JUL-2004 (first entry)
XX
DE Epigene construct PfCTL/HTL(N) protein.
XX
KW hepatotropic; virucide; antiinflammatory; anti-HIV; cytostatic; epitope;
KW cytotoxic T lymphocyte; CTL; helper T lymphocyte; HTL; immune response;
KW hepatitis B virus; HBV; HIV; cancer; lymphoma.
XX
OS Unidentified.
XX
PN WO2004031210-A2.
XX
PD 15-APR-2004.
XX
PF 03-OCT-2003; 2003WO-US031303.
XX
PR 03-OCT-2002; 2002US-0415463P.
PR 22-OCT-2002; 2002US-0419973P.
XX
PA (EPIM-) EPIMMUNE INC.
PA (GEMV) GENENCOR INT INC.
XX
PI Sette A, Chesnut R, Newman MJ, Livingston BD, Babe LM, Chen Y;
PI Deyoung LM, Huang MTF, Power SD;
XX
DR WPI; 2004-330143/30.
DR N-PSDB; ADO24116.
XX
PT New multi-epitope polynucleotides encoding cytotoxic T lymphocyte and/or
PT helper T lymphocyte epitopes, useful for inducing or stimulating an
PT antiviral or anticancer immune response, especially against hepatitis B
PT virus.
XX
PS Disclosure; SEQ ID NO 265; 401pp; English.
XX
CC The invention relates to a multi-epitope polynucleotide (I) encoding
CC cytotoxic T lymphocyte (CTL) and/or helper T lymphocyte (HTL) epitopes.
CC The composition and methods are useful for inducing or stimulating an
CC immune response against hepatitis B virus (HBV), other viruses (e.g.
CC HIV), or against cancers such as prostate cancer or lymphoma. This
CC sequence corresponds to an example of an epigene construct encoded
CC protein used in the invention.
XX
SQ Sequence 456 AA;

Query Match 17.0%; Score 94; DB 8; Length 456;
Best Local Similarity 85.7%; Pred. No. 0.12;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 30 PAIELPSENERGYYIPHQSSL 50
| ||||||||||||||||||
Db 119 PGPGLPSENERGYYIPHQSSL 139

```
RESULT 11
AAW85415
ID   AAW85415 standard; peptide; 15 AA.
XX
AC   AAW85415;
XX
DT   16-FEB-1999  (first entry)
XX
DE   Helper T-cell class II peptide derived from LSA1 protein.
XX
KW   Helper T-cell peptide; human leucocyte antigen; HLA; DR4w4; DR1; DR7;
KW   cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;
KW   acquired immune deficiency syndrome; malaria; cancer;
KW   allograft rejection; allergy; Lyme disease; hepatitis;
KW   post-streptococcal endocarditis; glomerulonephritis;
KW   food hypersensitivity.
XX
OS   Synthetic.
OS   Plasmodium falciparum.
XX
PN   WO9832456-A1.
XX
PD   30-JUL-1998.
XX
PF   23-JAN-1998;   98WO-US001373.
XX
PR   23-JAN-1997;   97US-0036713P.
PR   07-FEB-1997;   97US-0037432P.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 1998-427679/36.
XX
PT   Composition containing peptide that induces cytotoxic T lymphocyte
PT   response, and helper peptide - can bind to human leucocyte antigen
PT   alleles, used to treat or prevent cancers, parasitic infections and
PT   autoimmune disease.
XX
PS   Disclosure; Page 42; 51pp; English.
XX
CC   AAW85284-451 represent helper T-cell class II peptides, which can bind to
CC   the human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides are
CC   used in the course of the invention. The specification describes peptides
CC   that that induce a cytotoxic T lymphocyte (CTL) response, and T-helper
CC   peptides, that are used together to generate a CTL response for the
CC   treatment or prevention of viral, fungal, bacterial or parasitic
CC   infections (e.g. hepatitis, acquired immune deficiency syndrome or
CC   malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate
CC   cancer or condyloma acuminatum). Helper T-cell peptides may be used alone
CC   to induce a helper T cell response, e.g. in cases of autoimmune disease,
CC   allograft rejection, allergy, Lyme disease, hepatitis, post-streptococcal
CC   endocarditis, glomerulonephritis and food hypersensitivity
XX
SQ   Sequence 15 AA;

  Query Match             15.3%;  Score 85;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 RGYYIPHQSSLPQDN 54
              |||||||||||||||
Db          1 RGYYIPHQSSLPQDN 15


RESULT 12
AAW85172
ID   AAW85172 standard; peptide; 15 AA.
XX
AC   AAW85172;
XX
DT   16-FEB-1999  (first entry)
XX
DE   Helper T-cell peptide derived from a LSA1 protein.
XX
KW   Helper T-cell peptide; human leucocyte antigen; HLA; DR4w4; DR1; DR7;
KW   cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;
KW   acquired immune deficiency syndrome; malaria; cancer;
KW   allograft rejection; allergy; Lyme disease; hepatitis;
KW   post-streptococcal endocarditis; glomerulonephritis;
KW   food hypersensitivity.
XX
OS   Synthetic.
OS   Plasmodium falciparum.
XX
PN   WO9832456-A1.
XX
PD   30-JUL-1998.
XX
PF   23-JAN-1998;   98WO-US001373.
XX
PR   23-JAN-1997;   97US-0036713P.
PR   07-FEB-1997;   97US-0037432P.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 1998-427679/36.
XX
PT   Composition containing peptide that induces cytotoxic T lymphocyte
PT   response, and helper peptide - can bind to human leucocyte antigen
PT   alleles, used to treat or prevent cancers, parasitic infections and
PT   autoimmune disease.
XX
PS   Claim 11; Page 37; 51pp; English.
XX
CC   AAW85138-283 represent helper T-cell peptides, which can bind to the
CC   human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides are used
CC   in the course of the invention. The specification describes peptides that
CC   that induce a cytotoxic T lymphocyte (CTL) response, and T-helper
CC   peptides, that are used together to generate a CTL response for the
CC   treatment or prevention of viral, fungal, bacterial or parasitic
CC   infections (e.g. hepatitis, acquired immune deficiency syndrome or
CC   malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate
CC   cancer or condyloma acuminatum). Helper T-cell peptides may be used alone
CC   to induce a helper T cell response, e.g. in cases of autoimmune disease,
CC   allograft rejection, allergy, Lyme disease, hepatitis, post-streptococcal
CC   endocarditis, glomerulonephritis and food hypersensitivity
XX
SQ   Sequence 15 AA;

  Query Match             15.3%;  Score 85;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 RGYYIPHQSSLPQDN 54
              |||||||||||||||
Db          1 RGYYIPHQSSLPQDN 15


RESULT 13
ADW36112
ID   ADW36112 standard; peptide; 15 AA.
XX
AC   ADW36112;
XX
DT   10-MAR-2005  (first entry)
XX
DE   HLA binding epitope #6862.
XX
KW   Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW   MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;
KW   viral disease; cancer.
XX
```

OS Unidentified.
XX
PN WO2003040165-A2.
XX
PD 15-MAY-2003.
XX
PF 18-OCT-2001; 2001WO-US051650.
XX
PR 19-OCT-2000; 2000US-0242350P.
PR 20-APR-2001; 2001US-0285624P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S;
XX
DR WPI; 2003-441519/41.
XX
PT New composition comprising at least one peptide having allele-specific
PT binding motifs for HLA, useful for preventing, treating or diagnosing
PT viral diseases and cancer.
XX
PS Claim 1; Page 52-379; 382pp; English.
XX
CC The invention relates to a composition comprising at least one peptide
CC having an isolated, prepared epitope selected from any of the sequences
CC from 30 lists given in the specification. Also disclosed is a method for
CC inducing a cytotoxic T cell response against a pre-selected antigen in a
CC patient expressing a specific MHC class I allele by contacting cytotoxic
CC T cells from the patient with the composition cited above. The
CC composition comprises an epitope that is joined by an amino acid linker.
CC The epitope is admixed or joined to a CTL or HTL epitope. The epitope is
CC bound to an HLA molecule on the antigen-presenting cell, where when an A2
CC -restricted cytotoxic lymphocyte (CTL) is present, a receptor of the CTL
CC binds to a complex of the HLA molecule and the epitope. Specifically
CC claimed are peptides having allele-specific binding motifs for HLA. The
CC compositions and methods are useful for preventing, treating or
CC diagnosing viral diseases and cancer. The peptide epitopes are useful as
CC diagnostic agents for evaluating immune responses, for making antibodies
CC and for evaluating efficacy of a vaccine. Sequences given in ADW29251-
CC ADW37745 represent epitopes of the invention as given in Tables 2-31.
XX
SQ Sequence 15 AA;

Query Match 15.3%; Score 85; DB 7; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 RGYYIPHQSSLPQDN 54
|||||||||||||||
Db 1 RGYYIPHQSSLPQDN 15

RESULT 14
ADW35943
ID ADW35943 standard; peptide; 15 AA.
XX
AC ADW35943;
XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #6693.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;
KW viral disease; cancer.
XX
OS Unidentified.
XX
PN WO2003040165-A2.
XX
PD 15-MAY-2003.
XX
PF 18-OCT-2001; 2001WO-US051650.
XX
PR 19-OCT-2000; 2000US-0242350P.
PR 20-APR-2001; 2001US-0285624P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S;
XX
DR WPI; 2003-441519/41.
XX
PT New composition comprising at least one peptide having allele-specific
PT binding motifs for HLA, useful for preventing, treating or diagnosing
PT viral diseases and cancer.
XX
PS Claim 1; Page 52-379; 382pp; English.
XX
CC The invention relates to a composition comprising at least one peptide
CC having an isolated, prepared epitope selected from any of the sequences
CC from 30 lists given in the specification. Also disclosed is a method for
CC inducing a cytotoxic T cell response against a pre-selected antigen in a
CC patient expressing a specific MHC class I allele by contacting cytotoxic
CC T cells from the patient with the composition cited above. The
CC composition comprises an epitope that is joined by an amino acid linker.
CC The epitope is admixed or joined to a CTL or HTL epitope. The epitope is
CC bound to an HLA molecule on the antigen-presenting cell, where when an A2
CC -restricted cytotoxic lymphocyte (CTL) is present, a receptor of the CTL
CC binds to a complex of the HLA molecule and the epitope. Specifically
CC claimed are peptides having allele-specific binding motifs for HLA. The
CC compositions and methods are useful for preventing, treating or
CC diagnosing viral diseases and cancer. The peptide epitopes are useful as
CC diagnostic agents for evaluating immune responses, for making antibodies
CC and for evaluating efficacy of a vaccine. Sequences given in ADW29251-
CC ADW37745 represent epitopes of the invention as given in Tables 2-31.
XX
SQ Sequence 15 AA;

Query Match 15.3%; Score 85; DB 7; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 RGYYIPHQSSLPQDN 54
|||||||||||||||
Db 1 RGYYIPHQSSLPQDN 15

RESULT 15
AEA26643
ID AEA26643 standard; protein; 829 AA.
XX
AC AEA26643;
XX
DT 28-JUL-2005 (first entry)
XX
DE Stress tolerant plant-related transcription factor protein SeqID484.
XX
KW transcription factor; transgenic plant; agriculture; drought resistance;
KW stress tolerance.
XX
OS Arabidopsis thaliana.
XX
PN WO2005047516-A2.
XX
PD 26-MAY-2005.
XX
PF 12-NOV-2004; 2004WO-US037584.
XX
PR 13-NOV-2003; 2003US-00714887.
PR 05-DEC-2003; 2003US-0527658P.
PR 05-FEB-2004; 2004US-0542928P.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
PI Heard JE, Riechmann JL, Creelman RA, Ratcliffe OJ, Canales RD;

PI Repetti P, Kumimoto RW, Gutterson NI, Reuber TL, Pineda O;
PI Sherman BK, Morrison TA, Keddie JS, Jiang C, Century KS, Adam L;
PI Zhang JZ, Hempel FD, Libby JM;
XX
DR WPI; 2005-372386/38.
DR N-PSDB; AEA26642.
XX
PT New transgenic plants for producing commercially or agriculturally useful
PT plants having improved tolerance to drought, shade and low nitrogen
PT conditions.
XX
PS Claim 46; SEQ ID NO 484; 407pp; English.
XX
CC This invention relates to a novel plant transcription factor
CC polypeptides, the DNA sequences which encode them and their use in
CC creating transgenic plants. The transgenic plant and methods are useful
CC for producing commercially or agriculturally useful plants having
CC improved tolerance to drought, shade and low nitrogen conditions when
CC compared to wild-type reference plants. The present sequence is that of a
CC plant transcription factor protein which was used during the development
CC of the transgenic plants of the invention.
XX
SQ Sequence 829 AA;

Query Match 14.8%; Score 82; DB 9; Length 829;
Best Local Similarity 27.7%; Pred. No. 6.2;
Matches 33; Conservative 21; Mismatches 39; Indels 26; Gaps 6;

Qy 2 KADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSR-D 60
Db 57 KDGIKKQMTPPAEEGN------GQKFEDAPVETPNEDKKG---DGDASLPKEDESSSKQD 107

Qy 61 SKEISIIEKTNRE--------SITTNVEGRRDIHK------GHLEEKKDGSIKPEQKED 105
Db 108 NQEEKKEEKTKEEFTPSSETKSETEGGEDQKDDSKSENGGGGDLDEKKD--LKDNSDEE 164

RESULT 16
AAG98701
ID AAG98701 standard; protein; 231 AA.
XX
AC AAG98701;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717 ORF5 protein, SEQ:194.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033547.
XX
PR 14-DEC-1999; 99US-00461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
DR WPI; 2001-390297/41.
DR N-PSDB; AAH84226, AAH84231.
XX
PT Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders.
XX
PS Claim 1; Fig 9E; 325pp; English.
XX
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC protective ORFs. The cell death protective cDNA clones are able to
CC prevent, delay or reverse progression through the apoptotic or necrotic
CC pathways when injected into a cell predisposed to or undergoing cell
CC death. The cell death protective nucleic acids and polypeptides can be
CC used in the diagnosis and treatment of disorders associated with cell
CC death, and to screen for compounds which modulate their activity or
CC expression. Such modulators, preferably a small organic molecule, an
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC cell death-related diseases. Such diseases include those associated with
CC the central nervous system including psychiatric or neurological
CC disorders, especially ischaemia-related conditions such as strokes, and
CC also includes neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, or Parkinson's disease. The modulators may also be
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC ; eye conditions such as diabetic retinopathy or macular degeneration;
CC hypertension; myocardial infarction; atherosclerosis; respiratory
CC conditions such as asthma or chronic obstructive pulmonary disease;
CC neoplastic conditions such as cancers or benign tumours; blood cell
CC conditions such as anaemia; gastrointestinal conditions such as gastritis
CC or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC conditions such as dermatitis or urticaria; or immune system disorders
CC such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC additionally be used to generate animal models of cell death-associated
CC disorders. The present sequence represents a cell death protective
CC polypeptide
XX
SQ Sequence 231 AA;

Query Match 14.7%; Score 81.5; DB 4; Length 231;
Best Local Similarity 29.2%; Pred. No. 1.3;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 78 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 137

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 138 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 183

RESULT 17
AAG98700
ID AAG98700 standard; protein; 232 AA.
XX
AC AAG98700;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717 ORF4 protein, SEQ:192.

XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033547.
XX
PR 14-DEC-1999; 99US-00461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
DR WPI; 2001-390297/41.
DR N-PSDB; AAH84226, AAH84230.
XX
PT Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders.
XX
PS Claim 1; Fig 9D; 325pp; English.
XX
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC protective ORFs. The cell death protective cDNA clones are able to
CC prevent, delay or reverse progression through the apoptotic or necrotic
CC pathways when injected into a cell predisposed to or undergoing cell
CC death. The cell death protective nucleic acids and polypeptides can be
CC used in the diagnosis and treatment of disorders associated with cell
CC death, and to screen for compounds which modulate their activity or
CC expression. Such modulators, preferably a small organic molecule, an
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC cell death-related diseases. Such diseases include those associated with
CC the central nervous system including psychiatric or neurological
CC disorders, especially ischaemia-related conditions such as strokes, and
CC also includes neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, or Parkinson's disease. The modulators may also be
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC ; eye conditions such as diabetic retinopathy or macular degeneration;
CC hypertension; myocardial infarction; atherosclerosis; respiratory
CC conditions such as asthma or chronic obstructive pulmonary disease;
CC neoplastic conditions such as cancers or benign tumours; blood cell
CC conditions such as anaemia; gastrointestinal conditions such as gastritis
CC or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC conditions such as dermatitis or urticaria; or immune system disorders
CC such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC additionally be used to generate animal models of cell death-associated
CC disorders. The present sequence represents a cell death protective
CC polypeptide
XX
SQ Sequence 232 AA;

Query Match 14.7%; Score 81.5; DB 4; Length 232;
Best Local Similarity 29.2%; Pred. No. 1.3;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 79 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 138

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 139 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 184

RESULT 18
AAG98699
ID AAG98699 standard; protein; 238 AA.
XX
AC AAG98699;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717 ORF3 protein, SEQ:190.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033547.
XX
PR 14-DEC-1999; 99US-00461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
DR WPI; 2001-390297/41.
DR N-PSDB; AAH84226, AAH84229.
XX
PT Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders.
XX
PS Claim 1; Fig 9C; 325pp; English.
XX
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences

```
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective
CC   polypeptide
XX
SQ   Sequence 238 AA;

  Query Match             14.7%;  Score 81.5;  DB 4;  Length 238;
  Best Local Similarity   29.2%;  Pred. No. 1.3;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    |  : ||:|     :      :| :||  |
Db         85 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 144

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
                ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        145 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 190


RESULT 19
AAG98698
ID   AAG98698 standard; protein; 257 AA.
XX
AC   AAG98698;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human cell death protective cDNA clone CNI-00717 ORF2 protein, SEQ:188.
XX
KW   Cell death protective; apoptosis; necrosis; human; drug screening;
KW   cell death-associated disorder; central nervous system disorder;
KW   psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW   stroke; cerebral infarction; ischaemic encephalopathy;
KW   neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW   Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW   vascular disease; opthalmological disorder; diabetic retinopathy;
KW   macular degeneration; hypertension; myocardial infarction;
KW   atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW   chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW   benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW   ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW   glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW   Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW   urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   N-PSDB; AAH84226, AAH84228.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 1; Fig 9B; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective
CC   polypeptide
XX
SQ   Sequence 257 AA;

  Query Match             14.7%;  Score 81.5;  DB 4;  Length 257;
  Best Local Similarity   29.2%;  Pred. No. 1.5;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    |  : ||:|     :      :| :||  |
Db        104 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 163

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
                ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        164 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 209


RESULT 20
```

```
AAG98697
ID   AAG98697 standard; protein; 272 AA.
XX
AC   AAG98697;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human cell death protective cDNA clone CNI-00717 ORF1 protein, SEQ:186.
XX
KW   Cell death protective; apoptosis; necrosis; human; drug screening;
KW   cell death-associated disorder; central nervous system disorder;
KW   psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW   stroke; cerebral infarction; ischaemic encephalopathy;
KW   neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW   Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW   vascular disease; opthalmological disorder; diabetic retinopathy;
KW   macular degeneration; hypertension; myocardial infarction;
KW   atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW   chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW   benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW   ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW   glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW   Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW   urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   N-PSDB; AAH84226, AAH84227.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 1; Fig 9A; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective
CC   polypeptide
XX
SQ   Sequence 272 AA;

  Query Match             14.7%;  Score 81.5;  DB 4;  Length 272;
  Best Local Similarity   29.2%;  Pred. No. 1.6;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    | : ||:|     :      :| :||  |
Db        119 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 178

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |       ||  :| ||  :| ::||
Db        179 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 224


RESULT 21
ABP26206
ID   ABP26206 standard; protein; 379 AA.
XX
AC   ABP26206;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 1588.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB004789.
XX
PR   27-OCT-2000; 2000GB-00026333.
PR   24-NOV-2000; 2000GB-00028727.
PR   07-MAR-2001; 2001GB-00005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Y RosI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN66837.
XX
PT   New Streptococcus protein for the treatment or prevention of infection or
PT   disease caused by Streptococcus bacteria, such as meningitis, and for
PT   detecting a compound that binds to the protein.
XX
PS   Claim 1; Page 3309; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
```

```
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins
XX
SQ   Sequence 379 AA;

  Query Match             14.7%;  Score 81.5;  DB 5;  Length 379;
  Best Local Similarity   28.6%;  Pred. No. 2.5;
  Matches   22;  Conservative   18;  Mismatches   20;  Indels   17;  Gaps    3;

Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::         |:|:::|:|  :    |:       |:  |  |::||
Db        239 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 298

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        299 EE------EPEKQEDES 309


RESULT 22
ADU69665
ID   ADU69665 standard; protein; 412 AA.
XX
AC   ADU69665;
XX
DT   10-FEB-2005  (first entry)
XX
DE   S agalactiae hyperimmune serum reactive antigen seqid 360.
XX
KW   immune stimulation; antigen; bacterial surface display;
KW   hyperimmune serum reactive antigen; vaccine; bacterial infection;
KW   antibacterial; infection.
XX
OS   Streptococcus agalactiae.
XX
PN   WO2004099242-A2.
XX
PD   18-NOV-2004.
XX
PF   06-MAY-2004; 2004WO-EP004856.
XX
PR   07-MAY-2003; 2003EP-00450112.
PR   28-NOV-2003; 2003EP-00450266.
XX
PA   (INTE-) INTERCELL AG.
XX
PI   Meinke A,  Nagy E,  Hanner M,  Horky M,  Kallenda S,  Prustomersky S;
XX
DR   WPI; 2004-821662/81.
DR   N-PSDB; ADU69448.
XX
PT   New nucleic acid molecule encoding a hyperimmune serum reactive antigen,
PT   useful for the manufacture of a vaccine against Streptococcus agalactiae
PT   infection.
XX
PS   Claim 12; SEQ ID NO 360; 221pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule encoding a
CC   hyperimmune serum reactive antigen or its fragment. Also described are: a
CC   vector comprising the nucleic acid molecule; a host cell comprising the
CC   vector; a hyperimmune serum-reactive antigen comprising a sequence
CC   encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467
CC   or 812 amino acids; fragments of hyperimmune serum-reactive antigens
CC   consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;
CC   a process for producing a Streptococcus agalactiae hyperimmune serum
CC   reactive antigen; a process for producing a cell that expresses a S.
CC   agalactiae hyperimmune serum reactive antigen; a pharmaceutical
CC   composition, especially a vaccine, comprising the hyperimmune serum-
CC   reactive antigen or nucleic acid molecule; an antibody that binds at
CC   least to a selective part of the hyperimmune serum-reactive antigen; a
CC   hybridoma cell line, which produces the antibody; a method for producing
CC   the antibody; an antagonist that binds to the hyperimmune serum-reactive
CC   antigen; a method for identifying an antagonist capable of binding to the
CC   hyperimmune serum-reactive antigen; a method for identifying an
CC   antagonist capable of reducing or inhibiting the interaction activity of
CC   a hyperimmune serum-reactive antigen to its interaction partner; a
CC   process for in vitro diagnosing a disease related to expression of the
CC   hyperimmune serum-reactive antigen; and a process for in vitro diagnosis
CC   of a bacterial infection, especially a S. agalactiae infection. The
CC   hyperimmune serum reactive antigen is useful for isolating, purifying
CC   and/or identifying an interaction partner of the hyperimmune serum
CC   reactive antigen. The hyperimmune serum reactive antigen is useful for
CC   generating a peptide binding to the hyperimmune serum reactive antigen,
CC   where the peptide comprises anticalines, or for the manufacture of a
CC   functional nucleic acid comprising aptamers or spiegelmers. The nucleic
CC   acid molecule is useful for the manufacture of a functional ribonucleic
CC   acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic
CC   acid molecule, hyperimmune serum-reactive antigen or antibody is useful
CC   for the manufacture of a vaccine against S. agalactiae infection. This is
CC   the amino acid sequence of a Streptococcus agalactiae hyperimmune serum
CC   reactive antigen.
XX
SQ   Sequence 412 AA;

  Query Match             14.7%;  Score 81.5;  DB 8;  Length 412;
  Best Local Similarity   28.6%;  Pred. No. 2.8;
  Matches   22;  Conservative   18;  Mismatches   20;  Indels   17;  Gaps    3;

Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::         |:|:::|:|  :    |:       |:  |  |::||
Db        272 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 331

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        332 EE------EPEKQEDES 342


RESULT 23
ADV89658
ID   ADV89658 standard; protein; 412 AA.
XX
AC   ADV89658;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Streptococcus agalactiae protein sequence, SEQ ID 2052.
XX
KW   Antibacterial; Vaccine; bacterial infection.
XX
OS   Streptococcus agalactiae.
XX
PN   FR2824074-A1.
XX
PD   31-OCT-2002.
XX
PF   26-APR-2001; 2001FR-00005642.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P,  Rusniok C,  Chevalier F,  Frangeul L,  Lalioui L;
PI   Zouine M,  Couve E,  Buchrieser C,  Poyart C,  Trieu CP,  Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
```

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 6; SEQ ID NO 2052; 2687pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
CC agalactiae involved in the synthesis of amino acids, cell membranes,
CC intermediate (central) metabolism, energetic metabolism, fatty acid and
CC phospholipid metabolism, nucleotide metabolism including purines,
CC pyrimidines and/or nucleosides, regulatory functions, replication,
CC transcription, translation, protein transport, adaptation to atypical
CC conditions, sensitivity to medicines and/or analogues, functions related
CC to transposons, biosynthesis of cofactors, prosthetic groups and
CC transporters, cell membrane proteins and cellular machinery. (I) are
CC useful for the detection and/or amplification of nucleic acids.
CC Pharmaceutical composition comprising (I) or (II) are useful for
CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
CC equivalent for the present basic patent FR2824074A1. WO200292818A2
CC contains 6617 sequence whereas the present patent only contains 2344
CC sequences.
XX
SQ Sequence 412 AA;

Query Match 14.7%; Score 81.5; DB 8; Length 412;
Best Local Similarity 28.6%; Pred. No. 2.8;
Matches 22; Conservative 18; Mismatches 20; Indels 17; Gaps 3;

```
Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::         |:|:::|:|  :    |:       |:  |  |::||
Db        272 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 331

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        332 EE------EPEKQEDES 342
```

RESULT 24
ADV83060
ID ADV83060 standard; protein; 412 AA.
XX
AC ADV83060;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae protein, SEQ ID 4201.
XX
KW Antibacterial; vaccine; bacterial infection.
XX
OS Streptococcus agalactiae.
XX
PN WO200292818-A2.
XX
PD 21-NOV-2002.
XX
PF 26-APR-2002; 2002WO-IB003059.
XX
PR 26-APR-2001; 2001FR-00005642.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR WPI; 2004-101891/11.
XX
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 6; SEQ ID NO 4201; 439pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.
XX
SQ Sequence 412 AA;

Query Match 14.7%; Score 81.5; DB 8; Length 412;
Best Local Similarity 28.6%; Pred. No. 2.8;
Matches 22; Conservative 18; Mismatches 20; Indels 17; Gaps 3;

```
Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::         |:|:::|:|  :    |:       |:  |  |::||
Db        272 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 331

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        332 EE------EPEKQEDES 342
```

RESULT 25
ADV80911
ID ADV80911 standard; protein; 412 AA.
XX
AC ADV80911;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae protein, SEQ ID 2052.
XX
KW Antibacterial; vaccine; bacterial infection.
XX
OS Streptococcus agalactiae.
XX
PN WO200292818-A2.
XX
PD 21-NOV-2002.
XX
PF 26-APR-2002; 2002WO-IB003059.
XX
PR 26-APR-2001; 2001FR-00005642.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR WPI; 2004-101891/11.
XX
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 6; SEQ ID NO 2052; 439pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The

```
CC   nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC   synthesis of amino acids, cell membranes, intermediate (central)
CC   metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC   nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC   regulatory functions, replication, transcription, translation, protein
CC   transport, adaptation to atypical conditions, sensitivity to medicines
CC   and/or analogues, functions related to transposons, biosynthesis of
CC   cofactors, prosthetic groups and transporters, cell membrane proteins and
CC   cellular machinery. (I) are useful for the detection and/or amplification
CC   of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC   useful for treatment of a bacterial S. agalactiae infection. The complete
CC   genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC   present patent is an equivalent for the basic patent FR2824074A1, which
CC   contains only 2344 sequences.
XX
SQ   Sequence 412 AA;

  Query Match             14.7%;  Score 81.5;  DB 8;  Length 412;
  Best Local Similarity   28.6%;  Pred. No. 2.8;
  Matches   22;  Conservative   18;  Mismatches   20;  Indels   17;  Gaps    3;

Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::          |:|:::|:|  :    |:       |:  |  |::||
Db        272 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 331

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        332 EE------EPEKQEDES 342


RESULT 26
AAY91310
ID   AAY91310 standard; protein; 485 AA.
XX
AC   AAY91310;
XX
DT   30-MAY-2000  (first entry)
XX
DE   Group B Streptococcus protein sequence SEQ ID NO:38.
XX
KW   Group B Streptococcus; Streptococcus agalactiae; protein antigen;
KW   vaccine; screening; immunogen; detection; diagnosis; infection; antibody;
KW   affibody; antibacterial.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200006736-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB002444.
XX
PR   27-JUL-1998;   98GB-00016335.
PR   19-MAR-1999;   99US-0125163P.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Le Page RWF,  Wells JM,  Hanniffy SB;
XX
DR   WPI; 2000-195299/17.
XX
PT   New Group B Streptococcus protein, useful as vaccine, for diagnosis of
PT   Streptococcal infections and for screening of antibodies or affibodies.
XX
PS   Claim 1; Fig 1; 123pp; English.
XX
CC   AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given in
CC   AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also
CC   known as Streptococcus agalactiae. The GBS polynucleotides and
CC   polypeptides have antibacterial activity. Immunogenic compositions
CC   comprising GBS polynucleotides or polypeptides can be used as vaccines
CC   and for the treatment or prophylaxis of GBS infection. The
CC   polynucleotides and polypeptides can also be used in the detection of GBS
CC   and for screening DNA encoding bacterial cell envelope associated or
CC   secreted antigens in gram positive bacteria. AAA05873 to AAA05941
CC   represent primers used in the exemplification of the present invention
XX
SQ   Sequence 485 AA;

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 485;
  Best Local Similarity   28.6%;  Pred. No. 3.4;
  Matches   22;  Conservative   18;  Mismatches   20;  Indels   17;  Gaps    3;

Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::          |:|:::|:|  :    |:       |:  |  |::||
Db        375 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 434

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        435 EE------EPEKQEDES 445


RESULT 27
ABG22380
ID   ABG22380 standard; protein; 2447 AA.
XX
AC   ABG22380;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #22371.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS86567.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 52739; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
```

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2447 AA;

Query Match 14.7%; Score 81.5; DB 4; Length 2447;
Best Local Similarity 30.9%; Pred. No. 30;
Matches 30; Conservative 13; Mismatches 31; Indels 23; Gaps 5;

Qy 16 GDILAEDLYG--RLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSK---EISIIEKT 70
Db 1676 GPLLASD-FGPDRIFIPA----------PVVSSMPERASGSSSGEKGGYEIFVKVKD 1721

Qy 71 NRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
Db 1722 THEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 1754

RESULT 28
ADX79752
ID ADX79752 standard; protein; 294 AA.
XX
AC ADX79752;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 49118.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
DR WPI; 2004-180133/17.
XX
PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 49118; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 294 AA;

Query Match 14.6%; Score 81; DB 8; Length 294;
Best Local Similarity 23.2%; Pred. No. 2;
Matches 26; Conservative 24; Mismatches 32; Indels 30; Gaps 5;

Qy 4 DTKKNLE---------RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDN 54
Db 178 DTAMEIEAPVEVSGKKAKKKHTAVVTDDM-------AVDKPTETTNGDAL-------EDH 223

Qy 55 RGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
Db 224 KSEKKKRK------KEKRKLDTENDQAMDDGANG-VESEQDGAVKKKKKKDK 268

RESULT 29
ADZ72255
ID ADZ72255 standard; protein; 371 AA.
XX
AC ADZ72255;
XX
DT 14-JUL-2005 (first entry)
XX
DE Plasmodium falciparum MSP-3-like protein, MSP-3-2 SEQ ID NO: 4.
XX
KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;
KW infection; merozoite surface protein 3-like protein; MSP-3-2; antigen.
XX
OS Plasmodium falciparum.
XX
PN EP1526178-A1.
XX
PD 27-APR-2005.
XX
PF 24-OCT-2003; 2003EP-00292673.
XX
PR 24-OCT-2003; 2003EP-00292673.
XX
PA (INSP) INST PASTEUR.
XX
PI Druilhe P;
XX
DR WPI; 2005-323987/34.
DR N-PSDB; ADZ72254.
XX
PT Novel MSP-3-like family genes located on chromosome 10 of Plasmodium
PT falciparum, which encode proteins useful for preparing vaccine
PT compositions against malaria.
XX
PS Disclosure; SEQ ID NO 4; 137pp; English.
XX
CC The present invention relates to the protection against malaria. More
CC particularly, the invention pertains to a family of MSP-3 (merozoite
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium
CC falciparum, highly conserved in P. falciparum strains, simultaneously
CC expressed in P. falciparum at the erythrocytic stages and encoding
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at

```
CC   their N-terminal extremity and which are located at the merozoite
CC   surface. The characterization of this gene family enables the definition
CC   of immunogenic and vaccine compositions against P. falciparum. The
CC   present sequence is the P. falciparum MSP-3-2 protein.
XX
SQ   Sequence 371 AA;

  Query Match             14.5%;  Score 80.5;  DB 9;  Length 371;
  Best Local Similarity   30.2%;  Pred. No. 3.1;
  Matches   29;  Conservative   16;  Mismatches   38;  Indels   13;  Gaps    4;

Qy         17 DILAEDLYGRLEIPAIE---LPSENERGYYIPHQSSLPQDNRGNSRDSK---EISIIEKT 70
              |   | |  :::||: :   :| |||:    |      :||:    | :   |    |
Db        199 DKKTEYLLEQIKIPSWDRNNIPDENEQVIEDPQ-----EDNKDEDEDEETETENLETEDD 253

Qy         71 NRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              | | |  | |   || :  :|||::   | |::| |
Db        254 NNEEIEENEED--DIDEESVEEKEEEEEKKEEEEKK 287


RESULT 30
ABU24404
ID   ABU24404 standard; protein; 903 AA.
XX
AC   ABU24404;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #9931.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Clostridium botulinum.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA28274.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 52328; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 903 AA;

  Query Match             14.5%;  Score 80.5;  DB 6;  Length 903;
  Best Local Similarity   25.0%;  Pred. No. 10;
  Matches   28;  Conservative   24;  Mismatches   43;  Indels   17;  Gaps    5;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQS-SLPQD-------NRGN 57
                :|| :  ||| ::  |    : ::| : :  : |    :|::|  | ::
Db        272 EKNKQSIKEHNELREET---KKKLPKVNIELKEETKKQVPNKSLELDEEKDKQVFKEHNE 328

Qy         58 SRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGS--IKPEQKEDKS 107
               |: |   |||: |     :  |  : : ||::| ||: |  |  |  | |:
Db        329 LRNEKSKQIIEEHNE----SKDEKSKQVSKGNIELKKEKSKQIIKEHNESKN 376


Search completed: December 15, 2005, 21:44:47
Job time : 195 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:41:17 ; Search time 47 Seconds
(without alignments)
188.219 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	107	2	US-08-098-327E-19	Sequence 19, Appl
2	554	100.0	107	2	US-08-462-625-19	Sequence 19, Appl
3	554	100.0	117	2	US-08-098-327E-20	Sequence 20, Appl
4	554	100.0	117	2	US-08-462-625-20	Sequence 20, Appl
5	554	100.0	316	2	US-08-098-327E-31	Sequence 31, Appl
6	554	100.0	316	2	US-08-462-625-31	Sequence 31, Appl
7	154	27.8	31	2	US-08-098-327E-23	Sequence 23, Appl
8	154	27.8	31	2	US-08-462-625-23	Sequence 23, Appl
9	141	25.5	27	2	US-08-098-327E-21	Sequence 21, Appl
10	141	25.5	27	2	US-08-462-625-21	Sequence 21, Appl
11	97	17.5	709	2	US-09-248-796A-19045	Sequence 19045, A
12	85	15.3	15	2	US-09-009-953-74	Sequence 74, Appl
13	81.5	14.7	231	2	US-09-461-697-194	Sequence 194, App
14	81.5	14.7	232	2	US-09-461-697-192	Sequence 192, App
15	81.5	14.7	238	2	US-09-461-697-190	Sequence 190, App
16	81.5	14.7	257	2	US-09-461-697-188	Sequence 188, App
17	81.5	14.7	272	2	US-09-461-697-186	Sequence 186, App
18	79.5	14.4	1761	2	US-09-561-709B-1	Sequence 1, Appli
19	79	14.3	538	2	US-09-252-991A-23551	Sequence 23551, A
20	78	14.1	447	2	US-09-710-279-2900	Sequence 2900, Ap
21	78	14.1	450	2	US-09-710-279-3226	Sequence 3226, Ap
22	77.5	14.0	474	2	US-09-543-681A-6752	Sequence 6752, Ap
23	76.5	13.8	765	2	US-09-538-092-906	Sequence 906, App
24	76.5	13.8	765	2	US-09-882-274-2	Sequence 2, Appli
25	76.5	13.8	1085	1	US-08-431-080-28	Sequence 28, Appl
26	76.5	13.8	1085	1	US-08-938-534-28	Sequence 28, Appl
27	76.5	13.8	1085	2	US-09-345-294-28	Sequence 28, Appl
28	76	13.7	817	2	US-09-248-796A-20276	Sequence 20276, A
29	75.5	13.6	150	2	US-09-395-689-2	Sequence 2, Appli
30	75.5	13.6	300	2	US-09-395-689-1	Sequence 1, Appli
31	75.5	13.6	765	1	US-08-663-112-2	Sequence 2, Appli
32	73.5	13.3	547	2	US-09-252-991A-32044	Sequence 32044, A
33	73	13.2	566	2	US-09-270-767-43747	Sequence 43747, A
34	73	13.2	584	2	US-09-949-016-8587	Sequence 8587, Ap
35	73	13.2	585	2	US-09-949-016-6627	Sequence 6627, Ap
36	72.5	13.1	956	2	US-09-949-016-8159	Sequence 8159, Ap
37	72.5	13.1	1523	2	US-09-538-092-955	Sequence 955, App
38	72	13.0	495	2	US-09-248-796A-21845	Sequence 21845, A
39	71.5	12.9	714	2	US-09-949-016-10296	Sequence 10296, A
40	71.5	12.9	808	2	US-10-104-047-3401	Sequence 3401, Ap
41	71	12.8	783	6	5231168-2	Patent No. 5231168
42	70	12.6	159	2	US-09-640-211A-809	Sequence 809, App
43	70	12.6	449	1	US-08-839-008-2	Sequence 2, Appli
44	70	12.6	458	2	US-09-949-016-7238	Sequence 7238, Ap
45	70	12.6	754	2	US-09-976-594-375	Sequence 375, App

ALIGNMENTS

```
RESULT 1
US-08-098-327E-19
; Sequence 19, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-19

  Query Match 100.0%;  Score 554;  DB 2;  Length 107;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.9e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 2
US-08-462-625-19
; Sequence 19, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-19

  Query Match             100.0%;  Score 554;  DB 2;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 3
US-08-098-327E-20
; Sequence 20, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-20

  Query Match             100.0%;  Score 554;  DB 2;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 70

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         71 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 117


RESULT 4
US-08-462-625-20
; Sequence 20, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
```

```
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-20

  Query Match             100.0%;  Score 554;  DB 2;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 70

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         71 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 117


RESULT 5
US-08-098-327E-31
; Sequence 31, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-31

  Query Match             100.0%;  Score 554;  DB 2;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 269

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        270 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 316


RESULT 6
US-08-462-625-31
; Sequence 31, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
```

```
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-31

  Query Match             100.0%;  Score 554;  DB 2;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-58;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 269

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        270 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 316


RESULT 7
US-08-098-327E-23
; Sequence 23, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-23

  Query Match             27.8%;  Score 154;  DB 2;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-11;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 87
              |||||||||||||||||||||||||||||||
Db          1 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 31


RESULT 8
US-08-462-625-23
; Sequence 23, Application US/08462625
; Patent No. 6319502
; GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
```

```
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-23

  Query Match             27.8%;  Score 154;  DB 2;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-11;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 87
              ||||||||||||||||||||||||||||||
Db          1 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 31


RESULT 9
US-08-098-327E-21
; Sequence 21, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-21

  Query Match             25.5%;  Score 141;  DB 2;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-10;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIP 30
              |||||||||||||||||||||||||||
Db          1 DTKKNLERKKEHGDILAEDLYGRLEIP 27


RESULT 10
US-08-462-625-21
; Sequence 21, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-21

  Query Match             25.5%;  Score 141;  DB 2;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-10;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIP 30
              |||||||||||||||||||||||||||
Db          1 DTKKNLERKKEHGDILAEDLYGRLEIP 27


RESULT 11
US-09-248-796A-19045
; Sequence 19045, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
```

```
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19045
;   LENGTH: 709
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-19045

  Query Match             17.5%;  Score 97;  DB 2;  Length 709;
  Best Local Similarity   24.7%;  Pred. No. 0.0095;
  Matches   37;  Conservative   29;  Mismatches   36;  Indels   48;  Gaps    9;

Qy          1 RKADTKKNLERKKEHGDI---------LAEDLYGRLEIPA-------IELPSENERGYYI 44
              :: ||  |:: |||   |         : |||  ::| |:       :|:||::|  : |
Db        462 KEVDTDSNIQEKKELFQIKEPVKDEKQIKEDLV-KVETPSKDEGPAKVEVPSKDETPFKI 520

Qy         45 PHQSSLPQDNRGNSRDSK----------EISIIEKTNRESI----------TTNVEGRRD 84
               :  |   || | ::| |          :  :: | ::: |            : ||::|
Db        521 QNIDS-TTDN-GKAKDEKDNVSKVDSNVDTKVVTKDDQDGIKKVEEKPVEKKDDDEGKKD 578

Qy         85 -----IHKGHLEEKKDGSIKPE----QKED 105
                   :|| : || ||  :       ||:|
Db        579 NEKVTVHKDNEEEDKDKGVNEAKDKVQKDD 608


RESULT 12
US-09-009-953-74
; Sequence 74, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 74:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-009-953-74

  Query Match             15.3%;  Score 85;  DB 2;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.0011;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 RGYYIPHQSSLPQDN 54
              |||||||||||||||
Db          1 RGYYIPHQSSLPQDN 15


RESULT 13
US-09-461-697-194
; Sequence 194, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 194
;   LENGTH: 231
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-461-697-194

  Query Match             14.7%;  Score 81.5;  DB 2;  Length 231;
  Best Local Similarity   29.2%;  Pred. No. 0.14;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    |  : ||:|     :      :| :||  |
Db         78 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 137

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | | |       ||  :| ||  :| ::||
Db        138 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 183


RESULT 14
US-09-461-697-192
; Sequence 192, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 192
;   LENGTH: 232
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-461-697-192
```

Query Match 14.7%; Score 81.5; DB 2; Length 232;
Best Local Similarity 29.2%; Pred. No. 0.14;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 79 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 138

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 139 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 184

RESULT 15
US-09-461-697-190
; Sequence 190, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-190

Query Match 14.7%; Score 81.5; DB 2; Length 238;
Best Local Similarity 29.2%; Pred. No. 0.15;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 85 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 144

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 145 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 190

RESULT 16
US-09-461-697-188
; Sequence 188, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-188

Query Match 14.7%; Score 81.5; DB 2; Length 257;
Best Local Similarity 29.2%; Pred. No. 0.16;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 104 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 163

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 164 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 209

RESULT 17
US-09-461-697-186
; Sequence 186, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-186

Query Match 14.7%; Score 81.5; DB 2; Length 272;
Best Local Similarity 29.2%; Pred. No. 0.17;
Matches 31; Conservative 16; Mismatches 54; Indels 5; Gaps 3;

Qy 4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
Db 119 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 178

Qy 61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
Db 179 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 224

RESULT 18
US-09-561-709B-1
; Sequence 1, Application US/09561709B
; Patent No. 6682911
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert
; APPLICANT: Champliaud, Marie-France
; APPLICANT: Olson, Pamela
; APPLICANT: Koch, Manuel
; APPLICANT: Brunken, William
; TITLE OF INVENTION: LAMININS AND USES THEREOF
; FILE REFERENCE: 10287-060001
; CURRENT APPLICATION NUMBER: US/09/561,709B
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 09/168,949
; PRIOR FILING DATE: 1998-10-09

```
;  PRIOR APPLICATION NUMBER: US 60/061,609
;  PRIOR FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1761
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-561-709B-1

  Query Match             14.4%;  Score 79.5;  DB 2;  Length 1761;
  Best Local Similarity   29.1%;  Pred. No. 4.2;
  Matches   32;  Conservative   15;  Mismatches   34;  Indels   29;  Gaps    5;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIPAIEL------PSENERGYYIPHQSSLPQDNRGN 57
              | |  :|| |   |:| |||   ::: :  |       ||| : ||  | ||
Db       1270 DLKDTIERAKNEADLLLEDLQEEIDLQSSVLNASIADSSENIKKYY--HISS-------- 1319

Qy         58 SRDSKEISIIEKTNRESITTNVEGRRDI--------HKGHLEEKKDGSIK 99
                | |  | | ::  :| |:   | |:         ||:|  ::   ||
Db       1320 ---SAEKKINETSS--TINTSANTRNDLLTILDTLTSKGNLSLERLKQIK 1364


RESULT 19
US-09-252-991A-23551
; Sequence 23551, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 23551
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23551

  Query Match             14.3%;  Score 79;  DB 2;  Length 538;
  Best Local Similarity   26.8%;  Pred. No. 0.91;
  Matches   34;  Conservative   16;  Mismatches   41;  Indels   36;  Gaps    5;

Qy          5 TKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI----PH--QSSLPQDNR--- 55
              | :|||:    |    |||: || : ::|:|   |||: |     |  | :  |  |
Db        370 THRNLEKMVAEGS-FREDLFYRLNVLSLEVPPLRERGHDILLLARHFMQQACAQIQRPVC 428

Qy         56 ---------------GNSRDSK----------EISIIEKTNRESITTNVEGRRDIHKGH 89
                            || |  :          | |:::  : |   | |  : |   |
Db        429 RLAPGTYPALLSNRWPGNVRQLQNVIFRAAAICESSLVDIGDLEIAGTAVARQNDGEVGS 488

Qy         90 LEEKKDG 96
              |||  :|
Db        489 LEEAVEG 495


RESULT 20
US-09-710-279-2900
; Sequence 2900, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2900
;   LENGTH: 447
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-710-279-2900

  Query Match             14.1%;  Score 78;  DB 2;  Length 447;
  Best Local Similarity   24.8%;  Pred. No. 0.92;
  Matches   27;  Conservative   20;  Mismatches   36;  Indels   26;  Gaps    4;

Qy          4 DTKKNLE-----RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNS 58
              :| :||:     :|||: ::  | |   :|  |              | |   |:|
Db         28 ETARNLKADIDAQKKEYEEL--EQLSKEIEASA------------PKQDEPPKDEGAEV 72

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                |:|: :  |::  :      ||  |      ||| | : ||: | :::
Db         73 EDNKDGNSGEESENKPSDDEPEGTSD------EEKPDDAPKPDDKPEET 115


RESULT 21
US-09-710-279-3226
; Sequence 3226, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3226
;   LENGTH: 450
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-710-279-3226

  Query Match             14.1%;  Score 78;  DB 2;  Length 450;
  Best Local Similarity   24.8%;  Pred. No. 0.93;
  Matches   27;  Conservative   20;  Mismatches   36;  Indels   26;  Gaps    4;

Qy          4 DTKKNLE-----RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNS 58
              :| :||:     :|||: ::  | |   :|  |              | |   |:|
Db         28 ETARNLKADIDAQKKEYEEL--EQLSKEIEASA------------PKQDEPPKDEGAEV 72

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                |:|: :  |::  :      ||  |      ||| | : ||: | :::
Db         73 EDNKDGNSGEESENKPSDDEPEGTSD------EEKPDDAPKPDDKPEET 115


RESULT 22
US-09-543-681A-6752
; Sequence 6752, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
```

```
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6752
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-6752

  Query Match             14.0%;  Score 77.5;  DB 2;  Length 474;
  Best Local Similarity   30.6%;  Pred. No. 1.1;
  Matches   26;  Conservative   13;  Mismatches   29;  Indels   17;  Gaps    3;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSK 62
              | | ::||::   ||   |||| || :  |:||   :|   ||              |
Db        280 AATHQDLEKRVMEGD-FREDLYHRLNVIRIQLPPLRDRTEDIP--------------SL 323

Qy         63 EISIIEKTNRE-SITTNVEGRRDIH 86
                ::|| ||  : | |  :: :|
Db        324 ARYFLQKTARELGVETKVLHQQSLH 348


RESULT 23
US-09-538-092-906
; Sequence 906, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 906
;   LENGTH: 765
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P11387
US-09-538-092-906

  Query Match             13.8%;  Score 76.5;  DB 2;  Length 765;
  Best Local Similarity   26.9%;  Pred. No. 3;
  Matches   29;  Conservative   18;  Mismatches   40;  Indels   21;  Gaps    5;

Qy          6 KKNLERKKEHGD--ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-----PQDNRGNS 58
              |:  |: :  ||  |  |   |    | |:   |:: ||::| :  :     |:|
Db         88 KRKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDD-GYFVPPKEDIKPLKRPRDEDDAD 146

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                |:|   | | :|             |  |||::|| :|  : :||
Db        147 YKPKKIK-TEDTKKEK-----------KRKLEEEEDGKLKKPKNKDK 181


RESULT 24
US-09-882-274-2
; Sequence 2, Application US/09882274
; Patent No. 6951720
; GENERAL INFORMATION:
;  APPLICANT: Burgin, Alex B.
;  APPLICANT:  Stewart, Lance J.
;  TITLE OF INVENTION: Use of Phosphorothiolate Polynucleotides
;  TITLE OF INVENTION:  In Ligating Nucleic Acids
;  FILE REFERENCE: P-RD 4806
;  CURRENT APPLICATION NUMBER: US/09/882,274
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 60/290,313
;  PRIOR FILING DATE: 2001-05-10
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 765
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-882-274-2

  Query Match             13.8%;  Score 76.5;  DB 2;  Length 765;
  Best Local Similarity   26.9%;  Pred. No. 3;
  Matches   29;  Conservative   18;  Mismatches   40;  Indels   21;  Gaps    5;

Qy          6 KKNLERKKEHGD--ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-----PQDNRGNS 58
              |:  |: :  ||  |  |   |    | |:   |:: ||::| :  :     |:|
Db         88 KRKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDD-GYFVPPKEDIKPLKRPRDEDDAD 146

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                |:|   | | :|             |  |||::|| :|  : :||
Db        147 YKPKKIK-TEDTKKEK-----------KRKLEEEEDGKLKKPKNKDK 181


RESULT 25
US-08-431-080-28
; Sequence 28, Application US/08431080
; Patent No. 5698686
;  GENERAL INFORMATION:
;    APPLICANT:  Gottschling, Daniel E.
;    APPLICANT:  Singer, Miriam S.
;    TITLE OF INVENTION:  Telomerase Compositions and Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TEXAS
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/431,080
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SN 08/326,781
;      FILING DATE:  October 20, 1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:155/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1085 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-431-080-28

  Query Match             13.8%;  Score 76.5;  DB 1;  Length 1085;
  Best Local Similarity   21.3%;  Pred. No. 4.8;
  Matches   30;  Conservative   15;  Mismatches   53;  Indels   43;  Gaps    4;

Qy          1 RKADTKKNLERKKE----------------------------------HGDI-LAEDLY 24
```

```
              :| : |: |::||:                                        | ||    |
Db        873 KKTEKKRQLKKKKKLLKIRKQRQKAIKEQETMNLQLGINGHEIIGNNNSHSDINTGTDFT 932

Qy         25 GRLEIPAIELPSE-NERGYYIPHQSSLPQDNRGNSRDSKEI------SIIEKTNRESITT 77
                |  ||||   |    ||| | |  |:      :| :       |: |   : ::
Db        933 TNENTPMNELPSHAPEDASLIPHNSDLAVDSNTRKNSTKSVGLDEIHEILGKDENDLLSV 992

Qy         78 NVEGRRDIHKGHLEEKKDGSI 98
                 |  :||: |  |  |
Db        993 GDINGYDAQEGHVIEDTDADI 1013


RESULT 26
US-08-938-534-28
; Sequence 28, Application US/08938534
; Patent No. 5916752
;  GENERAL INFORMATION:
;    APPLICANT:  Gottschling, Daniel E.
;    APPLICANT:  Singer, Miriam S.
;    TITLE OF INVENTION:  Telomerase Compositions and Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TEXAS
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,534
;      FILING DATE:  26-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/431,080
;      FILING DATE:
;      APPLICATION NUMBER:  SN 08/326,781
;      FILING DATE:  October 20, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:155/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1085 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-938-534-28

  Query Match             13.8%;  Score 76.5;  DB 1;  Length 1085;
  Best Local Similarity   21.3%;  Pred. No. 4.8;
  Matches   30;  Conservative   15;  Mismatches   53;  Indels   43;  Gaps    4;

Qy          1 RKADTKKNLERKKE----------------------------------HGDI-LAEDLY 24
              :| : |: |::||:                                  | ||    |
Db        873 KKTEKKRQLKKKKKLLKIRKQRQKAIKEQETMNLQLGINGHEIIGNNNSHSDINTGTDFT 932

Qy         25 GRLEIPAIELPSE-NERGYYIPHQSSLPQDNRGNSRDSKEI------SIIEKTNRESITT 77
                |  ||||   |    ||| | |  |:      :| :       |: |   : ::
Db        933 TNENTPMNELPSHAPEDASLIPHNSDLAVDSNTRKNSTKSVGLDEIHEILGKDENDLLSV 992

Qy         78 NVEGRRDIHKGHLEEKKDGSI 98
                 |  :||: |  |  |
Db        993 GDINGYDAQEGHVIEDTDADI 1013


RESULT 27
US-09-345-294-28
; Sequence 28, Application US/09345294
; Patent No. 6387619
;  GENERAL INFORMATION:
;       APPLICANT: Gottschling, Daniel E.
;                  Singer, Miriam S.
;       TITLE OF INVENTION: Telomerase Compositions and Methods
;       NUMBER OF SEQUENCES: 32
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Arnold, White & Durkee
;            STREET: P.O. Box 4433
;            CITY: Houston
;            STATE: TEXAS
;            COUNTRY: UNITED STATES OF AMERICA
;            ZIP: 77210
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/345,294
;            FILING DATE: 30-Jun-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/431,080
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Parker, David L.
;            REGISTRATION NUMBER: 32,165
;            REFERENCE/DOCKET NUMBER: ARCD:155/PAR
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (512) 418-3000
;            TELEFAX: (713) 789-2679
;            TELEX: 79-0924
;  INFORMATION FOR SEQ ID NO: 28:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1085 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-345-294-28

  Query Match             13.8%;  Score 76.5;  DB 2;  Length 1085;
  Best Local Similarity   21.3%;  Pred. No. 4.8;
  Matches   30;  Conservative   15;  Mismatches   53;  Indels   43;  Gaps    4;

Qy          1 RKADTKKNLERKKE----------------------------------HGDI-LAEDLY 24
              :| : |: |::||:                                  | ||    |
Db        873 KKTEKKRQLKKKKKLLKIRKQRQKAIKEQETMNLQLGINGHEIIGNNNSHSDINTGTDFT 932

Qy         25 GRLEIPAIELPSE-NERGYYIPHQSSLPQDNRGNSRDSKEI------SIIEKTNRESITT 77
                |  ||||   |    ||| | |  |:      :| :       |: |   : ::
Db        933 TNENTPMNELPSHAPEDASLIPHNSDLAVDSNTRKNSTKSVGLDEIHEILGKDENDLLSV 992

Qy         78 NVEGRRDIHKGHLEEKKDGSI 98
                 |  :||: |  |  |
Db        993 GDINGYDAQEGHVIEDTDADI 1013


RESULT 28
US-09-248-796A-20276
; Sequence 20276, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
```

```
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20276
;   LENGTH: 817
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-20276

  Query Match             13.7%;  Score 76;  DB 2;  Length 817;
  Best Local Similarity   18.1%;  Pred. No. 3.7;
  Matches   19;  Conservative   32;  Mismatches   52;  Indels    2;  Gaps    1;

Qy          2 KADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDS 61
              ||: |: :: :|:  ::  : |: : :    |   : :|   |        |:::   :
Db        338 KAEAKRQIDIEKKQKELELQQLHEKAKKEEHEAKEKEKRDIEIAKLERNAQNDKRKKAED 397

Qy         62 KEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              :: :  :    | : | ::  :|:   | |||:   :  |:||::
Db        398 EKYAEYQAGRNEEVQTKLQEYKDVEAKHAEEKQ--VLLDEKKENQ 440


RESULT 29
US-09-395-689-2
; Sequence 2, Application US/09395689
; Patent No. 6387684
; GENERAL INFORMATION:
;  APPLICANT: Hwang, Jaulang
;  APPLICANT: Hui, cho-Fat
;  APPLICANT: Chen, Tzong-Yueh
;  TITLE OF INVENTION: TOPOISOMERASE 1-MEDIATED DNA DELIVERY
;  FILE REFERENCE: 089191/024001
;  CURRENT APPLICATION NUMBER: US/09/395,689
;  CURRENT FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-395-689-2

  Query Match             13.6%;  Score 75.5;  DB 2;  Length 150;
  Best Local Similarity   26.9%;  Pred. No. 0.39;
  Matches   29;  Conservative   18;  Mismatches   40;  Indels   21;  Gaps    5;

Qy          6 KKNLERKKEHGD--ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-----PQDNRGNS 58
              |:  |: :  ||  |  |   |    | |:   |:: ||::| :  :     |:|
Db         38 KRKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDD-GYFVPPKEDIKPLKRPRDEDDVD 96

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                |:|   | | :|            |  |||::|| :|  : :||
Db         97 YKPKKIK-TEDTKKEK-----------KRKLEEEEDGKLKKPKNKDK 131


RESULT 30
US-09-395-689-1
; Sequence 1, Application US/09395689
; Patent No. 6387684
; GENERAL INFORMATION:
;  APPLICANT: Hwang, Jaulang
;  APPLICANT: Hui, cho-Fat
;  APPLICANT: Chen, Tzong-Yueh
;  TITLE OF INVENTION: TOPOISOMERASE 1-MEDIATED DNA DELIVERY
;  FILE REFERENCE: 089191/024001
;  CURRENT APPLICATION NUMBER: US/09/395,689
;  CURRENT FILING DATE: 1999-09-13
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 300
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-395-689-1

  Query Match             13.6%;  Score 75.5;  DB 2;  Length 300;
  Best Local Similarity   26.9%;  Pred. No. 1;
  Matches   29;  Conservative   18;  Mismatches   40;  Indels   21;  Gaps    5;

Qy          6 KKNLERKKEHGD--ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-----PQDNRGNS 58
              |:  |: :  ||  |  |   |    | |:   |:: ||::| :  :     |:|
Db         88 KRKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDD-GYFVPPKEDIKPLKRPRDEDDVD 146

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                |:|   | | :|            |  |||::|| :|  : :||
Db        147 YKPKKIK-TEDTKKEK-----------KRKLEEEEDGKLKKPKNKDK 181


Search completed: December 15, 2005, 21:50:19
Job time : 49 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:48:53 ; Search time 167 Seconds
(without alignments)
267.711 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	107	3	US-09-837-344-19	Sequence 19, Appl
2	554	100.0	107	3	US-09-900-963-19	Sequence 19, Appl
3	554	100.0	117	3	US-09-837-344-20	Sequence 20, Appl
4	554	100.0	117	3	US-09-900-963-20	Sequence 20, Appl
5	554	100.0	316	3	US-09-837-344-31	Sequence 31, Appl
6	554	100.0	316	3	US-09-900-963-31	Sequence 31, Appl
7	553	99.8	1909	5	US-10-732-923-3341	Sequence 3341, Ap
8	154	27.8	31	3	US-09-837-344-23	Sequence 23, Appl
9	154	27.8	31	3	US-09-900-963-23	Sequence 23, Appl
10	141	25.5	27	3	US-09-837-344-21	Sequence 21, Appl
11	141	25.5	27	3	US-09-900-963-21	Sequence 21, Appl
12	94	17.0	456	3	US-09-894-018-121	Sequence 121, App
13	94	17.0	456	5	US-10-474-960A-121	Sequence 121, App
14	85	15.3	15	4	US-10-103-395-74	Sequence 74, Appl
15	81.5	14.7	231	3	US-09-922-261-194	Sequence 194, App
16	81.5	14.7	232	3	US-09-922-261-192	Sequence 192, App
17	81.5	14.7	238	3	US-09-922-261-190	Sequence 190, App
18	81.5	14.7	257	3	US-09-922-261-188	Sequence 188, App
19	81.5	14.7	272	3	US-09-922-261-186	Sequence 186, App
20	81.5	14.7	485	3	US-09-769-736-72	Sequence 72, Appl
21	81.5	14.7	2447	5	US-10-450-763-52739	Sequence 52739, A
22	81	14.6	294	4	US-10-425-114-49118	Sequence 49118, A
23	81	14.6	396	4	US-10-424-599-168151	Sequence 168151,
24	81	14.6	542	4	US-10-424-599-234091	Sequence 234091,
25	80.5	14.5	903	4	US-10-282-122A-52328	Sequence 52328, A
26	80	14.4	575	5	US-10-739-930-6961	Sequence 6961, Ap
27	79.5	14.4	321	3	US-09-864-761-33718	Sequence 33718, A
28	79.5	14.4	361	3	US-09-864-761-33614	Sequence 33614, A
29	79.5	14.4	361	3	US-09-864-761-34141	Sequence 34141, A
30	79.5	14.4	1047	4	US-10-029-386-33522	Sequence 33522, A
31	79.5	14.4	1087	4	US-10-029-386-32407	Sequence 32407, A
32	79.5	14.4	1561	5	US-10-450-763-52738	Sequence 52738, A
33	79.5	14.4	1870	5	US-10-450-763-52737	Sequence 52737, A
34	79.5	14.4	2426	4	US-10-764-425-172	Sequence 172, App
35	78	14.1	257	4	US-10-425-114-58860	Sequence 58860, A
36	78	14.1	395	4	US-10-425-114-69744	Sequence 69744, A
37	78	14.1	399	4	US-10-425-115-226794	Sequence 226794,
38	78	14.1	419	4	US-10-425-115-226797	Sequence 226797,
39	77.5	14.0	874	6	US-11-097-143-18096	Sequence 18096, A
40	77	13.9	138	4	US-10-767-701-45736	Sequence 45736, A
41	77	13.9	224	4	US-10-425-115-221184	Sequence 221184,
42	77	13.9	232	4	US-10-425-115-221188	Sequence 221188,
43	77	13.9	550	4	US-10-369-493-5995	Sequence 5995, Ap
44	76.5	13.8	765	3	US-09-882-274-2	Sequence 2, Appli
45	76.5	13.8	765	4	US-10-408-765A-1149	Sequence 1149, Ap

ALIGNMENTS

```
RESULT 1
US-09-837-344-19
; Sequence 19, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 19:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 107 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 19:
```

```
US-09-837-344-19

  Query Match             100.0%;  Score 554;  DB 3;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-50;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 2
US-09-900-963-19
; Sequence 19, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 19:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 107 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-900-963-19

  Query Match             100.0%;  Score 554;  DB 3;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-50;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 3
US-09-837-344-20
; Sequence 20, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 20:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 117 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-837-344-20

  Query Match             100.0%;  Score 554;  DB 3;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-50;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 70

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         71 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 117


RESULT 4
US-09-900-963-20
; Sequence 20, Application US/09900963
```

```
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 20:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 117 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-900-963-20

  Query Match             100.0%;  Score 554;  DB 3;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-50;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 70

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         71 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 117


RESULT 5
US-09-837-344-31
; Sequence 31, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 31:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-837-344-31

  Query Match             100.0%;  Score 554;  DB 3;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-49;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 269

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        270 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 316


RESULT 6
US-09-900-963-31
; Sequence 31, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 31:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-900-963-31

  Query Match             100.0%;  Score 554;  DB 3;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-49;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 269

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        270 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 316


RESULT 7
US-10-732-923-3341
; Sequence 3341, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3341
;   LENGTH: 1909
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-732-923-3341

  Query Match             99.8%;  Score 553;  DB 5;  Length 1909;
  Best Local Similarity   99.1%;  Pred. No. 2e-48;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       1630 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 1689

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1690 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 1736


RESULT 8
US-09-837-344-23
; Sequence 23, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 23:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 31 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-837-344-23

  Query Match             27.8%;  Score 154;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-09;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 87
              |||||||||||||||||||||||||||||||
Db          1 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 31


RESULT 9
US-09-900-963-23
; Sequence 23, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
```

```
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Burns, Doane, Swecker & Mathis
;              STREET: P.O. Box 1404
;              CITY: Alexandria
;              STATE: Virginia
;              COUNTRY: United States
;              ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/900,963
;              FILING DATE: 10-Jul-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/098,327
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: McGowan, Malcolm K.
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-045
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 23:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 31 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-900-963-23

  Query Match             27.8%;  Score 154;  DB 3;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-09;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 87
              |||||||||||||||||||||||||||||||
Db          1 NSRDSKEISIIEKTNRESITTNVEGRRDIHK 31


RESULT 10
US-09-837-344-21
; Sequence 21, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;         APPLICANT: GUERIN-MARCHAND, Claudine
;                    DRUILHE, Pierre
;         TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                             HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                             OF STIMULATING THE T LYMPHOCYTES
;         NUMBER OF SEQUENCES: 46
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Burns, Doane, Swecker & Mathis
;              STREET: P.O. Box 1404
;              CITY: Alexandria
;              STATE: Virginia
;              COUNTRY: United States
;              ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/837,344
;              FILING DATE: 19-Apr-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/462,625
;              FILING DATE: <Unknown>
;              APPLICATION NUMBER: FR 91 01286
;              FILING DATE: 05-FEB-1991
;         ATTORNEY/AGENT INFORMATION:
;              NAME: McGowan, Malcolm K.
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-078
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 21:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 27 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-837-344-21

  Query Match             25.5%;  Score 141;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIP 30
              |||||||||||||||||||||||||||
Db          1 DTKKNLERKKEHGDILAEDLYGRLEIP 27


RESULT 11
US-09-900-963-21
; Sequence 21, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;         APPLICANT: GUERIN-MARCHAND, Claudine
;                    DRUILHE, Pierre
;         TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                             HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                             OF STIMULATING THE T LYMPHOCYTES
;         NUMBER OF SEQUENCES: 46
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Burns, Doane, Swecker & Mathis
;              STREET: P.O. Box 1404
;              CITY: Alexandria
;              STATE: Virginia
;              COUNTRY: United States
;              ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/900,963
;              FILING DATE: 10-Jul-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/098,327
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: McGowan, Malcolm K.
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-045
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
```

```
;   INFORMATION FOR SEQ ID NO: 21:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 27 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-900-963-21

  Query Match             25.5%;  Score 141;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIP 30
              |||||||||||||||||||||||||||
Db          1 DTKKNLERKKEHGDILAEDLYGRLEIP 27


RESULT 12
US-09-894-018-121
; Sequence 121, Application US/09894018
; Patent No. US20020119127A1
; GENERAL INFORMATION:
;  APPLICANT: EPIMMUNE, Inc.
;  APPLICANT:  Sette, Alessandro
;  APPLICANT:  Chestnut, Robert
;  APPLICANT:  Livingston, Brian
;  APPLICANT:  Baker, Denisw
;  APPLICANT:  Newman, Mark
;  APPLICANT:  Brown, David
;  TITLE OF INVENTION: METODS AND SYSTEM FOR OPTIMIZING
;  TITLE OF INVENTION:  MINIGENES AND PEPTIDES THEREBY
;  FILE REFERENCE: 39963-20033.00
;  CURRENT APPLICATION NUMBER: US/09/894,018
;  CURRENT FILING DATE:  2001-06-27
;  PRIOR APPLICATION NUMBER: PCT/US00/35568
;  PRIOR FILING DATE: 2000-12-28
;  PRIOR APPLICATION NUMBER: US 60/173,390
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: US 60/284,221
;  PRIOR FILING DATE: 2001-04-16
;  NUMBER OF SEQ ID NOS: 368
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 121
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PfCTL/HTL(N)
US-09-894-018-121

  Query Match             17.0%;  Score 94;  DB 3;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 0.39;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         30 PAIELPSENERGYYIPHQSSL 50
              |   ||||||||||||||||||
Db        119 PGPGLPSENERGYYIPHQSSL 139


RESULT 13
US-10-474-960A-121
; Sequence 121, Application US/10474960A
; Publication No. US20040248113A1
; GENERAL INFORMATION:
;  APPLICANT: Sette, Alessandro
;  APPLICANT:  Chesnut, Robert
;  APPLICANT:  Livingston, Brian
;  APPLICANT:  Baker, Denise
;  APPLICANT:  Newman, Mark
;  APPLICANT:  Brown, David
;  TITLE OF INVENTION: Methods and System for Optimizing Multi-epitope Nucleic
;  TITLE OF INVENTION:  Acid Constructs and Peptides Encoded Thereby
;  FILE REFERENCE: 2060.0320004
;  CURRENT APPLICATION NUMBER: US/10/474,960A
;  CURRENT FILING DATE:  2003-10-16
;  PRIOR APPLICATION NUMBER: PCT/US02/09877
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 09/894,018
;  PRIOR FILING DATE: 2001-06-27
;  PRIOR APPLICATION NUMBER: US 60/284,221
;  PRIOR FILING DATE: 2001-04-16
;  NUMBER OF SEQ ID NOS: 419
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 121
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PfCTL/HTL(N)
US-10-474-960A-121

  Query Match             17.0%;  Score 94;  DB 5;  Length 456;
  Best Local Similarity   85.7%;  Pred. No. 0.39;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         30 PAIELPSENERGYYIPHQSSL 50
              |   ||||||||||||||||||
Db        119 PGPGLPSENERGYYIPHQSSL 139


RESULT 14
US-10-103-395-74
; Sequence 74, Application US/10103395
; Publication No. US20020160019A1
; GENERAL INFORMATION:
;  APPLICANT: EPIMMUNE, Inc.
;  APPLICANT:  Sette, Alessandro
;  APPLICANT:  Sidney, John
;  APPLICANT:  Southwood, Scott
;  TITLE OF INVENTION: IDENTIFICATION OF BROADLY REACTIVE DR
;  TITLE OF INVENTION:  RESTRICTED EPITOPES
;  FILE REFERENCE: 39963-20016.01
;  CURRENT APPLICATION NUMBER: US/10/103,395
;  CURRENT FILING DATE:  2003-01-03
;  PRIOR APPLICATION NUMBER: US 09/009,953
;  PRIOR FILING DATE: 1998-01-21
;  PRIOR APPLICATION NUMBER: PCT/US98/01373
;  PRIOR FILING DATE: 1998-01-23
;  PRIOR APPLICATION NUMBER: US 60/036,713
;  PRIOR FILING DATE: 1997-01-23
;  PRIOR APPLICATION NUMBER: US 60/037,432
;  PRIOR FILING DATE: 1997-02-07
;  NUMBER OF SEQ ID NOS: 274
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 15
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-103-395-74

  Query Match             15.3%;  Score 85;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.041;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 RGYYIPHQSSLPQDN 54
              |||||||||||||||
Db          1 RGYYIPHQSSLPQDN 15


RESULT 15
US-09-922-261-194
```

```
; Sequence 194, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 194
;   LENGTH: 231
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-261-194

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 231;
  Best Local Similarity   29.2%;  Pred. No. 3.3;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    | : ||:|     :      :| :||  |
Db         78 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 137

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        138 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 183


RESULT 16
US-09-922-261-192
; Sequence 192, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 192
;   LENGTH: 232
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-261-192

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 232;
  Best Local Similarity   29.2%;  Pred. No. 3.4;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    | : ||:|     :      :| :||  |
Db         79 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 138

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        139 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 184


RESULT 17
US-09-922-261-190
; Sequence 190, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-261-190

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 238;
  Best Local Similarity   29.2%;  Pred. No. 3.5;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    | : ||:|     :      :| :||  |
Db         85 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 144

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        145 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 190


RESULT 18
US-09-922-261-188
; Sequence 188, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 188
;   LENGTH: 257
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-922-261-188

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 257;
  Best Local Similarity   29.2%;  Pred. No. 3.8;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    |  : ||:|     :      :| :||  |
Db        104 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 163

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        164 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 209


RESULT 19
US-09-922-261-186
; Sequence 186, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 186
;   LENGTH: 272
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-261-186

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 272;
  Best Local Similarity   29.2%;  Pred. No. 4.1;
  Matches   31;  Conservative   16;  Mismatches   54;  Indels    5;  Gaps    3;

Qy          4 DTKKNLE-RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ--SSLPQDNRGNSRD 60
              | |:: | : :| |:   ||   : |    |  : ||:|     :      :| :||  |
Db        119 DQKEDEEDQNEEKGEAGKEDKDEKGEEDGKEDKNGNEKGEDAKEKEDGKKGEDGKGNGED 178

Qy         61 SKEISIIEKTNRESITTNV--EGRRDIHKGHLEEKKDGSIKPEQKE 104
               ||    ||   :   | |  |      ||  :| ||  :| ::||
Db        179 GKEKGEDEKEEEDRKETGVGKENEDGKEKGDKKEGKDVKVKEDEKE 224


RESULT 20
US-09-769-736-72
; Sequence 72, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
;  APPLICANT: Microbial Technics Limited
;  APPLICANT:  Le Page, Richard WF
;  APPLICANT:  Wells, Jeremy M
;  APPLICANT:  Hanniffy, Sean B
;  TITLE OF INVENTION: Proteins
;  FILE REFERENCE: PWC/P21089wo
;  CURRENT APPLICATION NUMBER: US/09/769,736
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: GB 9816335.5
;  PRIOR FILING DATE: 1998-07-27
;  PRIOR APPLICATION NUMBER: US 60/125163
;  PRIOR FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
;   LENGTH: 485
;   TYPE: PRT
;   ORGANISM: Streptococcus agalactiae
US-09-769-736-72

  Query Match             14.7%;  Score 81.5;  DB 3;  Length 485;
  Best Local Similarity   28.6%;  Pred. No. 8.7;
  Matches   22;  Conservative   18;  Mismatches   20;  Indels   17;  Gaps    3;

Qy         42 YYIPHQSSLPQD---------NRGNSRDSKEISII--EKTNRESITTNVEGRRDIHKGHL 90
              ||: | : ||::          |:|:::|:|  :    |:       |:  |  |::||
Db        375 YYMEHPNELPKEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSD 434

Qy         91 EEKKDGSIKPEQKEDKS 107
              ||      :||::||:|
Db        435 EE------EPEKQEDES 445


RESULT 21
US-10-450-763-52739
; Sequence 52739, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 52739
;   LENGTH: 2447
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (2323)..(2357)
;   OTHER INFORMATION: 3 RETROVIRAL PROTEINASE domain identified by eMATRIX,
;   OTHER INFORMATION: accession number DM00892C, p-value=7.207e-28, raw score of 23.55
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (2385)..(2447)
;   OTHER INFORMATION: Double-stranded RNA binding motif domain identified by PFam,
;   OTHER INFORMATION: accession name dsrm, E-value=1.3e-18, PFam score of 75.2
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2447)
;   OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-52739

  Query Match             14.7%;  Score 81.5;  DB 5;  Length 2447;
  Best Local Similarity   30.9%;  Pred. No. 71;
  Matches   30;  Conservative   13;  Mismatches   31;  Indels   23;  Gaps    5;

Qy         16 GDILAEDLYG--RLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSK---EISIIEKT 70
              | :|| | :|  |: |||             |  ||:|:   |:|   |   || :  |
Db       1676 GPLLASD-FGPDRIFIPA------------PVVSSMPERASGSSSGEKGGYEIFVKVKD 1721

Qy         71 NRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                |    |    ::  ||  |:|:| |::   |  ||
Db       1722 THEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 1754
```

```
RESULT 22
US-10-425-114-49118
; Sequence 49118, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49118
;   LENGTH: 294
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700560825_FLI.pep
US-10-425-114-49118

  Query Match             14.6%;  Score 81;  DB 4;  Length 294;
  Best Local Similarity   23.2%;  Pred. No. 5.1;
  Matches   26;  Conservative   24;  Mismatches   32;  Indels   30;  Gaps    5;

Qy          4 DTKKNLE---------RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDN 54
              ||   :|          ||:|  :: :|:       |:: |:|   |  :       :|:
Db        178 DTAMEIEAPVEVSGKKAKKKHTAVVTDDM-------AVDKPTETTNGDAL-------EDH 223

Qy         55 RGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              :   :  |       |  |:  | | :   |   | :| ::||::| ::|:||
Db        224 KSEKKKRK------KEKRKLDTENDQAMDDGANG-VESEQDGAVKKKKKKDK 268


RESULT 23
US-10-424-599-168151
; Sequence 168151, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168151
;   LENGTH: 396
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(396)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_122856C.1.pep
US-10-424-599-168151

  Query Match             14.6%;  Score 81;  DB 4;  Length 396;
  Best Local Similarity   23.2%;  Pred. No. 7.6;
  Matches   26;  Conservative   24;  Mismatches   32;  Indels   30;  Gaps    5;

Qy          4 DTKKNLE---------RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDN 54
              ||   :|          ||:|  :: :|:       |:: |:|   |  :       :|:
Db        280 DTAMEIEAPVEVSGKKAKKKHTAVVTDDM-------AVDKPTETTNGDAL-------EDH 325

Qy         55 RGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              :   :  |       |  |:  | | :   |   | :| ::||::| ::|:||
Db        326 KSEKKKRK------KEKRKLDTENDQAMDDGANG-VESEQDGAVKKKKKKDK 370


RESULT 24
US-10-424-599-234091
; Sequence 234091, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234091
;   LENGTH: 542
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_53408C.1.pep
US-10-424-599-234091

  Query Match             14.6%;  Score 81;  DB 4;  Length 542;
  Best Local Similarity   23.8%;  Pred. No. 11;
  Matches   25;  Conservative   26;  Mismatches   30;  Indels   24;  Gaps    4;

Qy          2 KADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDS 61
              : : |: :| |:|:  : ||    : |: |:|:             :| :||    : ::
Db          6 EGEVKETIEGKQENEKVEAE----KPEVDAMEV------------ESGIPQKEGSDKKEK 49

Qy         62 KEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              |:::: |:           ||: :| |   ||| : | | |::  |
Db         50 KDVAMEEEDED-------EGKDNIDKSKEEEKAEDS-KGEKRSKK 86


RESULT 25
US-10-282-122A-52328
; Sequence 52328, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
```

```
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52328
;   LENGTH: 903
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-10-282-122A-52328

  Query Match             14.5%;  Score 80.5;  DB 4;  Length 903;
  Best Local Similarity   25.0%;  Pred. No. 25;
  Matches   28;  Conservative   24;  Mismatches   43;  Indels   17;  Gaps    5;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQS-SLPQD-------NRGN 57
              :|| :  ||| ::  |    : ::| : :  : |    :|::|  | ::
Db        272 EKNKQSIKEHNELREET---KKKLPKVNIELKEETKKQVPNKSLELDEEKDKQVFKEHNE 328

Qy         58 SRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGS--IKPEQKEDKS 107
               |: |   |||: |     :  | : : ||::| ||: |  |  |  | |:
Db        329 LRNEKSKQIIEEHNE----SKDEKSKQVSKGNIELKKEKSKQIIKEHNESKN 376


RESULT 26
US-10-739-930-6961
; Sequence 6961, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
;  FILE REFERENCE: 38-21(53377)B
;  CURRENT APPLICATION NUMBER: US/10/739,930
;  CURRENT FILING DATE:  2003-12-18
;  NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6961
;   LENGTH: 575
;   TYPE: PRT
;   ORGANISM: Brassica napus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: BRANA-23APR03-C293_1.p
US-10-739-930-6961

  Query Match             14.4%;  Score 80;  DB 5;  Length 575;
  Best Local Similarity   26.6%;  Pred. No. 16;
  Matches   29;  Conservative   16;  Mismatches   26;  Indels   38;  Gaps    5;

Qy         15 HGDILA-----EDLYGRLEIPAIELPSENERGYYIPHQSSL-----------------P 51
              |||: :     ||  | :  | | : |   ||| :|:|| :                  |
Db        162 HGDVASGGGAGEDKRGLVRTPTITVKS---RGYLVPNQSDVGIKGVRPPVLKPPPAKRPP 218

Qy         52 QDNRGNSRD-----SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKD 95
               |::|:| |     : |  |: : :  | : | |         |||: |
Db        219 IDHQGSSWDFLTHFAPEDGIVRRPSSSSTSDNGEET-------LEEEVD 260


RESULT 27
US-09-864-761-33718
; Sequence 33718, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33718
;   LENGTH: 321
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000046.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
;   OTHER INFORMATION: EST_HUMAN HIT: BE300756.1, EVALUE 2.00e-29
;   OTHER INFORMATION: EST_HUMAN HIT: BE835673.1, EVALUE 2.00e-29
;   OTHER INFORMATION: SWISSPROT HIT: P18583, EVALUE 3.00e-30
US-09-864-761-33718

  Query Match             14.4%;  Score 79.5;  DB 3;  Length 321;
  Best Local Similarity   26.6%;  Pred. No. 8.3;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   25;  Gaps    5;

Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |::||         | :|| |: ||            :|    |  ||:|:   ::
Db          8 KDMERLTSLRAGIEGPLLASDV-GR-----------DRSAASPVVSSMPERASESSSEE 54

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
```

```
              :|  ||  :  |   |    |    ::  ||  |:|:| |::   |  ||
Db          55 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 99


RESULT 28
US-09-864-761-33614
; Sequence 33614, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33614
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000190.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  15
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
;   OTHER INFORMATION: EST_HUMAN HIT: AW502403.1, EVALUE 5.00e-42
;   OTHER INFORMATION: SWISSPROT HIT: P18583, EVALUE 8.00e-43
US-09-864-761-33614

  Query Match             14.4%;  Score 79.5;  DB 3;  Length 361;
  Best Local Similarity   26.6%;  Pred. No. 9.6;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   25;  Gaps    5;

Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |::||          | :|| |: ||            :|    |  ||:|:   ::
Db         48 KDMERLTSLRAGIEGPLLASDV-GR------------DRSAASPVVSSMPERASESSSEE 94

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              :|  ||  :  |   |    |    ::  ||  |:|:| |::   |  ||
Db         95 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 139


RESULT 29
US-09-864-761-34141
; Sequence 34141, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34141
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   OTHER INFORMATION: MAP TO AP000114.1
;   OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL =   6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.6
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4
;   OTHER INFORMATION: EST_HUMAN HIT: AW502403.1, EVALUE 5.00e-42
;   OTHER INFORMATION: SWISSPROT HIT: P18583, EVALUE 8.00e-43
US-09-864-761-34141

  Query Match             14.4%;  Score 79.5;  DB 3;  Length 361;
  Best Local Similarity   26.6%;  Pred. No. 9.6;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   25;  Gaps    5;

Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |::||          | :|| |: ||             :|    |  ||:|:   ::
Db         48 KDMERLTSLRAGIEGPLLASDV-GR------------DRSAASPVVSSMPERASESSSEE 94

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              :|  || :  |   |    |    ::  ||  |:|:| |::   |  ||
Db         95 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 139


RESULT 30
US-10-029-386-33522
; Sequence 33522, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33522
;   LENGTH: 1047
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP001717.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
;   OTHER INFORMATION: SWISSPROT HIT: P18583, EVALUE 0.00e+00
US-10-029-386-33522

  Query Match             14.4%;  Score 79.5;  DB 4;  Length 1047;
  Best Local Similarity   26.6%;  Pred. No. 38;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   25;  Gaps    5;

Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |::||          | :|| |: ||             :|    |  ||:|:   ::
Db        734 KDMERLTSLRAGIEGPLLASDV-GR------------DRSAASPVVSSMPERASESSSEE 780

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              :|  || :  |   |    |    ::  ||  |:|:| |::   |  ||
Db        781 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 825


Search completed: December 15, 2005, 22:03:16
Job time : 169 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:49:34 ; Search time 13 Seconds
(without alignments)
55.432 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 51463 seqs, 6734788 residues

Total number of hits satisfying chosen parameters: 51463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79.5	14.4	2432	6	US-10-821-234-899	Sequence 899, App
2	78	14.1	447	6	US-10-793-626-2900	Sequence 2900, Ap
3	78	14.1	450	6	US-10-793-626-3226	Sequence 3226, Ap
4	71.5	12.9	212	6	US-10-467-657-5486	Sequence 5486, Ap
5	71.5	12.9	212	6	US-10-467-657-7404	Sequence 7404, Ap
6	71.5	12.9	677	6	US-10-982-545-12	Sequence 12, Appl
7	71	12.8	328	6	US-10-467-657-7710	Sequence 7710, Ap
8	68.5	12.4	693	7	US-11-196-475-72	Sequence 72, Appl
9	67	12.1	369	6	US-10-467-657-1336	Sequence 1336, Ap
10	65	11.7	360	6	US-10-878-556A-134	Sequence 134, App
11	65	11.7	360	7	US-11-186-284-115	Sequence 115, App
12	65	11.7	693	7	US-11-196-475-68	Sequence 68, Appl
13	65	11.7	708	7	US-11-196-475-76	Sequence 76, Appl
14	65	11.7	790	6	US-10-763-712A-75	Sequence 75, Appl
15	64.5	11.6	491	6	US-10-793-626-2808	Sequence 2808, Ap
16	64	11.6	658	6	US-10-467-657-4782	Sequence 4782, Ap
17	63.5	11.5	612	6	US-10-467-657-3988	Sequence 3988, Ap
18	62.5	11.3	505	6	US-10-821-234-1287	Sequence 1287, Ap
19	62.5	11.3	700	7	US-11-196-475-66	Sequence 66, Appl
20	62.5	11.3	795	6	US-10-770-726-49	Sequence 49, Appl
21	62	11.2	663	7	US-11-196-475-70	Sequence 70, Appl
22	62	11.2	663	7	US-11-196-475-78	Sequence 78, Appl
23	62	11.2	700	7	US-11-196-475-74	Sequence 74, Appl
24	61.5	11.1	451	6	US-10-467-657-366	Sequence 366, App
25	61.5	11.1	578	6	US-10-821-234-1039	Sequence 1039, Ap
26	61.5	11.1	858	6	US-10-645-441-25	Sequence 25, Appl
27	61.5	11.1	4384	6	US-10-821-234-1120	Sequence 1120, Ap
28	61	11.0	250	6	US-10-793-626-310	Sequence 310, App
29	61	11.0	250	6	US-10-793-626-2174	Sequence 2174, Ap
30	61	11.0	552	7	US-11-074-176-204	Sequence 204, App
31	61	11.0	574	6	US-10-467-657-2548	Sequence 2548, Ap
32	60.5	10.9	858	6	US-10-645-441-18	Sequence 18, Appl
33	60.5	10.9	858	6	US-10-645-441-23	Sequence 23, Appl
34	60	10.8	215	6	US-10-793-626-1650	Sequence 1650, Ap
35	60	10.8	440	7	US-11-108-172-1059	Sequence 1059, Ap
36	59.5	10.7	140	6	US-10-793-626-2268	Sequence 2268, Ap
37	59.5	10.7	140	6	US-10-793-626-2438	Sequence 2438, Ap
38	59.5	10.7	297	6	US-10-793-626-66	Sequence 66, Appl
39	59.5	10.7	297	6	US-10-793-626-746	Sequence 746, App
40	59.5	10.7	349	6	US-10-821-234-1387	Sequence 1387, Ap
41	59.5	10.7	387	6	US-10-485-517-132	Sequence 132, App
42	59.5	10.7	1189	6	US-10-821-234-1209	Sequence 1209, Ap
43	59	10.6	531	7	US-11-060-914-4	Sequence 4, Appli
44	59	10.6	1028	7	US-11-067-121-7	Sequence 7, Appli
45	58.5	10.6	793	7	US-11-060-914-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-821-234-899
; Sequence 899, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 899
; LENGTH: 2432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-899

Query Match 14.4%; Score 79.5; DB 6; Length 2432;
Best Local Similarity 26.6%; Pred. No. 3.4;
Matches 29; Conservative 17; Mismatches 38; Indels 25; Gaps 5;

```
Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |::||         | :|| |: ||            :|    |  ||:|:   ::
Db       1746 KDMERLTSLRAGIEGPLLASDV-GR-----------DRSAASPVVSSMPERASESSSEE 1792

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                :|  || :  |   |     |    ::  ||  |:|:| |::   |  ||
Db       1793 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 1837
```

RESULT 2
US-10-793-626-2900
; Sequence 2900, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

```
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2900
;   LENGTH: 447
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-2900

  Query Match             14.1%;  Score 78;  DB 6;  Length 447;
  Best Local Similarity   24.8%;  Pred. No. 0.58;
  Matches   27;  Conservative   20;  Mismatches   36;  Indels   26;  Gaps    4;

Qy          4 DTKKNLE-----RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNS 58
              :| :||:     :|||: ::  | |   :|  |             | |   |:|
Db         28 ETARNLKADIDAQKKEYEEL--EQLSKEIEASA------------PKQDEPPKDEGAEV 72

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                |:|: :  |::  :      ||  |      ||| | : ||: | :::
Db         73 EDNKDGNSGEESENKPSDDEPEGTSD------EEKPDDAPKPDDKPEET 115


RESULT 3
US-10-793-626-3226
; Sequence 3226, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3226
;   LENGTH: 450
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-3226

  Query Match             14.1%;  Score 78;  DB 6;  Length 450;
  Best Local Similarity   24.8%;  Pred. No. 0.58;
  Matches   27;  Conservative   20;  Mismatches   36;  Indels   26;  Gaps    4;

Qy          4 DTKKNLE-----RKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNS 58
              :| :||:     :|||: ::  | |   :|  |             | |   |:|
Db         28 ETARNLKADIDAQKKEYEEL--EQLSKEIEASA------------PKQDEPPKDEGAEV 72

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                |:|: :  |::  :      ||  |      ||| | : ||: | :::
Db         73 EDNKDGNSGEESENKPSDDEPEGTSD------EEKPDDAPKPDDKPEET 115


RESULT 4
US-10-467-657-5486
; Sequence 5486, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5486
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5486

  Query Match             12.9%;  Score 71.5;  DB 6;  Length 212;
  Best Local Similarity   23.2%;  Pred. No. 1;
  Matches   35;  Conservative   21;  Mismatches   38;  Indels   57;  Gaps    8;

Qy          4 DTKKNLERKKE----------------------------HGDILA--EDLYGRLEIPAI 32
              |:||  |:|::                            |  :||    :||
Db         56 DSKKLSEKKRDALAEMIKEQAVAWHVAASTPEEIASLNILHATMLAMKRAVYG------- 108

Qy         33 ELPSENER----GYYIPHQSSLPQDN--RGNSR--DSKEISIIEKTNRESITTNVEGRR- 83
               | :  |:     |  ||    :| :   :|:|:  :    |:: || |::    :  ||
Db        109 -LAARPEKIFIDGNRIPEHLGIPAEAVVKGDSKIIEISAASVLAKTARDAEMYALAQRRP 167

Qy         84 ----DIHKG-----HLEEKKDGSIKPEQKED 105
                  | |||     |||  |   : || : |
Db        168 QYGFDKHKGYGTKQHLEALKQYGVLPEHRRD 198


RESULT 5
US-10-467-657-7404
; Sequence 7404, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7404
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7404

  Query Match             12.9%;  Score 71.5;  DB 6;  Length 212;
  Best Local Similarity   23.2%;  Pred. No. 1;
  Matches   35;  Conservative   21;  Mismatches   38;  Indels   57;  Gaps    8;

Qy          4 DTKKNLERKKE----------------------------HGDILA--EDLYGRLEIPAI 32
              |:||  |:|::                            |  :||    :||
Db         56 DSKKLSEKKRDALAEMIKEQAVAWHVAASTPEEIASLNILHATMLAMKRAVYG------- 108

Qy         33 ELPSENER----GYYIPHQSSLPQDN--RGNSR--DSKEISIIEKTNRESITTNVEGRR- 83
               | :  |:     |  ||    :| :   :|:|:  :    |:: || |::    :  ||
Db        109 -LAARPEKIFIDGNRIPEHLGIPAEAVVKGDSKIIEISAASVLAKTARDAEMYALAQRRP 167

Qy         84 ----DIHKG-----HLEEKKDGSIKPEQKED 105
                  | |||     |||  |   : || : |
Db        168 QYGFDKHKGYGTKQHLEALKQYGVLPEHRRD 198


RESULT 6
US-10-982-545-12
; Sequence 12, Application US/10982545
; Publication No. US20050244890A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Davies, Huw Alun
;  APPLICANT:  McGuire, James
;  APPLICANT:  Simonsen, Anja Hviid
;  APPLICANT:  Blennow, Kaj
;  APPLICANT:  Podust, Vladimir
;  APPLICANT:  Ciphergen Biosystems, Inc.
;  TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
;  FILE REFERENCE: 016866-011550US
;  CURRENT APPLICATION NUMBER: US/10/982,545
;  CURRENT FILING DATE:  2004-11-06
;  PRIOR APPLICATION NUMBER: US 60/518,360
;  PRIOR FILING DATE: 2003-11-07
;  PRIOR APPLICATION NUMBER: US 60/526,753
;  PRIOR FILING DATE: 2003-12-02
;  PRIOR APPLICATION NUMBER: US 60/546,423
;  PRIOR FILING DATE: 2004-02-19
;  PRIOR APPLICATION NUMBER: US 60/547,250
;  PRIOR FILING DATE: 2004-02-23
;  PRIOR APPLICATION NUMBER: US 60/558,896
;  PRIOR FILING DATE: 2004-04-02
;  PRIOR APPLICATION NUMBER: US 60/572,617
;  PRIOR FILING DATE: 2004-05-18
;  PRIOR APPLICATION NUMBER: US 60/586,503
;  PRIOR FILING DATE: 2004-07-08
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 677
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Chromogranin B precursor
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: (1)..(20)
;   OTHER INFORMATION: signal peptide
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (21)..(677)
;   OTHER INFORMATION: Chromogranin B (Secretogranin I)
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (326)..(385)
;   OTHER INFORMATION: biomarker peptide 7258 Da, processed fragment of
;   OTHER INFORMATION: Chromogranin B
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (341)
;   OTHER INFORMATION: Xaa = sulfotyrosine
US-10-982-545-12

  Query Match             12.9%;  Score 71.5;  DB 6;  Length 677;
  Best Local Similarity   24.5%;  Pred. No. 4.4;
  Matches   27;  Conservative   15;  Mismatches   49;  Indels   19;  Gaps    3;

Qy          9 LERKKEHGDILAEDLYGRLEIPAI---------ELPSENERGYYIPHQSSLPQDNRGNSR 59
              |  :|:   |  |:|  | | |              | | |      | : ||:::|  |
Db        199 LNERKQASAIKKEELVARSETHAAGHSQEKTHSREKSSQESGEEAGSQENHPQESKGQPR 258

Qy         60 DSKEISIIEKTNRESITTNVEGRRDIHKGHL------EEKKDGSIKPEQK 103
                :|     :   |  |: |: ||   : |          : ||:  |:|
Db        259 SQEE----SEEGEEDATSEVDKRRTRPRHHHGRSRPDRSSQGGSLPSEEK 304


RESULT 7
US-10-467-657-7710
; Sequence 7710, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7710
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7710

  Query Match             12.8%;  Score 71;  DB 6;  Length 328;
  Best Local Similarity   30.9%;  Pred. No. 2;
  Matches   25;  Conservative   15;  Mismatches   35;  Indels    6;  Gaps    3;

Qy         25 GRLEIPAIELPSENERGYY-IPHQSSLPQDNR--GNSRDSKEISIIEKTNRESITTNVEG 81
              |::|   :|:  |  ||| :  :  :  :||  |  :     | | : : |     ||
Db        137 GQIE---MEIKVEQGRGYQSVSGRQVVRDENRQIGAIQLDASFSPISRVSFEVEPARVEQ 193

Qy         82 RRDIHKGHLEEKKDGSIKPEQ 102
              | |: |  |: : |||| ||:
Db        194 RTDLDKLVLDIETDGSIDPEE 214


RESULT 8
US-11-196-475-72
; Sequence 72, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-72

  Query Match             12.4%;  Score 68.5;  DB 7;  Length 693;
  Best Local Similarity   22.1%;  Pred. No. 9.3;
  Matches   27;  Conservative   23;  Mismatches   43;  Indels   29;  Gaps    4;

Qy          2 KADTKKNLERK-----KEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRG 56
              || :|: :: |     ||   :  ||             :|:     |    | :| :
Db        411 KASSKEKIKGKEKEIVKEKSKVSLGDL------------DNDETLMTPEDQKLSEDKKL 457

Qy         57 NSRDS-KEISIIEKTNRESITTNVEGRR----------DIHKGHLEEKKDGSIKPEQKED 105
              :|: : | :| ||: |  | : | |  :           ||    : : || :::  : :
Db        458 DSKKNLKPVSEIERVNEISKSNNNEVSKSSPLDKPSYSDIDSKEVVDNKDVNLQETKPQA 517
```

```
Qy        106 KS 107
              ||
Db        518 KS 519


RESULT 9
US-10-467-657-1336
; Sequence 1336, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1336
;   LENGTH: 369
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1336

  Query Match             12.1%;  Score 67;  DB 6;  Length 369;
  Best Local Similarity   27.4%;  Pred. No. 6;
  Matches   20;  Conservative   14;  Mismatches   29;  Indels   10;  Gaps    1;

Qy         30 PAIELPSENERGYYIPHQSSLPQDNRGNSRD----------SKEISIIEKTNRESITTNV 79
              |   : || :||  :|   :: ||| | ::|           :  :|| | ||   |::
Db        163 PGHTVRSEYKRGGGVPSLIAVYQDNSGKAKDIALSYAAANGGTKGGVIETTFREETETDL 222

Qy         80 EGRRDIHKGHLEE 92
                | : :  | : |
Db        223 FGEQAVLCGGVAE 235


RESULT 10
US-10-878-556A-134
; Sequence 134, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 134
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_hum/mk01_human
;   DATABASE ENTRY DATE: 1992-12-01
US-10-878-556A-134

  Query Match             11.7%;  Score 65;  DB 6;  Length 360;
  Best Local Similarity   23.9%;  Pred. No. 9.2;
  Matches   27;  Conservative   20;  Mismatches   34;  Indels   32;  Gaps    5;

Qy         16 GDILAEDLYGRLEIPA-------------IELPSENE---------RGYY--IPHQSSLP 51
              | |||| |  |   |               :  ||: :         | |   :||:: :|
Db        215 GCILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKARNYLLSLPHKNKVP 274

Qy         52 QDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKG----HLEEKKDGSIKP 100
              :    : ||| : :::|     :| |   | :: :     :||:  | | :|
Db        275 WNRLFPNADSKALDLLDK----MLTFNPHKRIEVEQALAHPYLEQYYDPSDEP 323


RESULT 11
US-11-186-284-115
; Sequence 115, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/11/186,284
;  CURRENT FILING DATE:  2005-07-21
;  PRIOR APPLICATION NUMBER: US/10/301,822
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-11-186-284-115

  Query Match             11.7%;  Score 65;  DB 7;  Length 360;
  Best Local Similarity   23.9%;  Pred. No. 9.2;
  Matches   27;  Conservative   20;  Mismatches   34;  Indels   32;  Gaps    5;

Qy         16 GDILAEDLYGRLEIPA-------------IELPSENE---------RGYY--IPHQSSLP 51
              | |||| |  |   |               :  ||: :         | |   :||:: :|
Db        215 GCILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKARNYLLSLPHKNKVP 274

Qy         52 QDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKG----HLEEKKDGSIKP 100
              :    : ||| : :::|     :| |   | :: :     :||:  | | :|
Db        275 WNRLFPNADSKALDLLDK----MLTFNPHKRIEVEQALAHPYLEQYYDPSDEP 323


RESULT 12
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
```

```
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-68

  Query Match             11.7%;  Score 65;  DB 7;  Length 693;
  Best Local Similarity   25.9%;  Pred. No. 21;
  Matches   29;  Conservative   17;  Mismatches   52;  Indels   14;  Gaps    4;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSK--- 62
              |||:   |   ||  : |     : |  | |||| |       :  |     :   |
Db        189 KKNILSGKVESDIDIDSLVTDKVVAA--LLSENEAGVNFARDITDIQGETHKADQDKIDI 246

Qy         63 EISIIEKTNRESITTNVEGRRD--------IHKGHLEEKKDGSIKPEQKEDK 106
              |:  : |::  :||  :|  ||          |:  :| : |   | ::: ||
Db        247 ELDNVHKSD-SNITETIENLRDQLEKATDEEHRKEIESQVDAKKKQKEELDK 297


RESULT 13
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 76
;   LENGTH: 708
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-76

  Query Match             11.7%;  Score 65;  DB 7;  Length 708;
  Best Local Similarity   19.8%;  Pred. No. 22;
  Matches   23;  Conservative   24;  Mismatches   33;  Indels   36;  Gaps    3;

Qy          1 RKADTKKNLERKKEHGDIL-AEDLYGRLEIPAIELPSE---------------------- 37
              :| : :| :|  ||   ::| ::|     :  |::|  |
Db        376 QKREIEKQIEIKKRDEELLKSKDGKVSKDYEALDLDRELSKASSKEKSKVKEEEITKGKS 435

Qy         38 --------NERGYYIPHQSSLPQDNRGNS-----RDSKEISIIEKTNRESITTNVE 80
                     |::   :|    ||:| : :|     :: | :| :|| :: | : | |
Db        436 RASLGDLNNDKNLMLPEDQKLPEDKKLDSKLDGKKEFKPVSEVEKLDKISKSNNNE 491


RESULT 14
US-10-763-712A-75
; Sequence 75, Application US/10763712A
; Publication No. US20050266541A1
; GENERAL INFORMATION:
;  APPLICANT: Solazyme, Inc.
;  APPLICANT:  Dillon, Harrison F.
;  TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen
;  TITLE OF INVENTION:  Production
;  FILE REFERENCE: H2042101-CIP
;  CURRENT APPLICATION NUMBER: US/10/763,712A
;  CURRENT FILING DATE:  2004-01-21
;  PRIOR APPLICATION NUMBER: US 10/287,750
;  PRIOR FILING DATE: 2002-11-04
;  PRIOR APPLICATION NUMBER: US 10/411,910
;  PRIOR FILING DATE: 2003-04-12
;  PRIOR APPLICATION NUMBER: US 60/500,032
;  PRIOR FILING DATE: 2003-09-03
;  NUMBER OF SEQ ID NOS: 184
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75
;   LENGTH: 790
;   TYPE: PRT
;   ORGANISM: Geobacter metallireducens
US-10-763-712A-75

  Query Match             11.7%;  Score 65;  DB 6;  Length 790;
  Best Local Similarity   23.2%;  Pred. No. 25;
  Matches   33;  Conservative   14;  Mismatches   43;  Indels   52;  Gaps    6;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIP-------HQSSLPQDNR 55
              | | ||||:     |    |||| || :  | :|    |   ||        ::|   : |
Db        608 AATNKNLEQGVREG-WFREDLYYRLNVFTITMPPLRSRVESIPILVHHFMDKASTKLNKR 666

Qy         56 ------------------GNSRDSKEI----------SIIEKTN---------RESITTN 78
                               || |: : :           :|   |          |   |
Db        667 MVGIEDRAVKALTKYPWPGNIREMQNVIERAAVLTHDGVIRVENFPLALSEGLEEGFATG 726

Qy         79 VE-------GRRDIHKGHLEEK 93
              ::         |: | | ||:|
Db        727 LDIHAASFRSEREQHMGKLEKK 748


RESULT 15
US-10-793-626-2808
; Sequence 2808, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2808
;   LENGTH: 491
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-2808

  Query Match             11.6%;  Score 64.5;  DB 6;  Length 491;
  Best Local Similarity   21.8%;  Pred. No. 15;
  Matches   22;  Conservative   22;  Mismatches   36;  Indels   21;  Gaps    6;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQD--NRGNS 58
              |  |  :   : |:  |:    :|  | :|         : ||:  : : ||    |:||
Db        367 RNVDDIEGYLKNKQSWDV---SMYSYLSVP------RGDTGYFF-NTAYLPDGALNKGNY 416

Qy         59 RDSKEISIIEKTN-------RESITTNV--EGRRDIHKGHL 90
              ::|   :|:: |       |  :|  :  | ::||   ::
Db        417 SNTKVTQLIKELNTTFGDKQRGQVTNEILNESKKDIPNSYI 457
```

```
RESULT 16
US-10-467-657-4782
; Sequence 4782, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 4782
;   LENGTH: 658
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-4782

  Query Match             11.6%;  Score 64;  DB 6;  Length 658;
  Best Local Similarity   19.5%;  Pred. No. 25;
  Matches   25;  Conservative   26;  Mismatches   51;  Indels   26;  Gaps    3;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDS- 61
              |||: ||      :   :  |: |    |   :||||:        |: |  |: ::|::
Db        326 ADTRANLTESVSNAGEVLHDITGVTPAP---MPSENDGENLFDSASNHPTGNKPDTRNAF 382

Qy         62 -----------------KEISIIEKTNRESITTNVEGRR-----DIHKGHLEEKKDGSIK 99
                               :: |:  : :| :: |  |  :     |:     |: : |::
Db        383 GSSGKTAPMPYQAARAPQQHSLSLRESRAAMDTYAELYKKTDDIDLELSQFEQARFGNMP 442

Qy        100 PEQKEDKS 107
                |    |:
Db        443 SETPAHKT 450


RESULT 17
US-10-467-657-3988
; Sequence 3988, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3988
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3988

  Query Match             11.5%;  Score 63.5;  DB 6;  Length 612;
  Best Local Similarity   25.5%;  Pred. No. 26;
  Matches   28;  Conservative   15;  Mismatches   42;  Indels   25;  Gaps    5;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |: | :   | || ||    : : |  |  |   | ||::|            :|| |
Db        110 REGDKESKAEHKKAHG----KPVSGSKEKNAKTQP-ENKQG-------KKGAKGQGNPRK 157

Qy         61 SKEISIIEKTNRESITTNVEGRRDIH-------KGHLEEKKDGSIKPEQK 103
                        |  ::::: | : | | :       | | |||   :   | |
Db        158 G------GKAEKDTVSANKKARSDKNGKAVKQDKKHTEEKNAKTDSDELK 201


RESULT 18
US-10-821-234-1287
; Sequence 1287, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1287
;   LENGTH: 505
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1287

  Query Match             11.3%;  Score 62.5;  DB 6;  Length 505;
  Best Local Similarity   28.7%;  Pred. No. 25;
  Matches   31;  Conservative   19;  Mismatches   35;  Indels   23;  Gaps    7;

Qy          7 KNLERK-KEHGDILAED---LYGRLEIPAIELPSENE-RG----YYIPHQSSL-PQDNRG 56
              |||| | || |: |::|    :  :::  | ::||  | ||    |: |    | |:   |
Db        410 KNLEPKYKELGEKLSKDPNIVIAKMDATANDVPSPYEVRGFPTIYFSPANKKLNPKKYEG 469

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKE 104
                 |  || ::   ||:    |           ::|:|    |  |::
Db        470 GRELSDFISYLQ---REATNPPV----------IQEEKPKKKKKAQED 504


RESULT 19
US-11-196-475-66
; Sequence 66, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-66
```

```
  Query Match             11.3%;  Score 62.5;  DB 7;  Length 700;
  Best Local Similarity   25.5%;  Pred. No. 38;
  Matches   28;  Conservative   19;  Mismatches   40;  Indels   23;  Gaps    5;

Qy         17 DILAEDLYGRLEIPAI-------ELPSENERGYYIPHQSSLPQDNRGNSRDSKEISI-IE 68
              |||: ::   ::| ::        | |||| |       :   | :| :  : :  | ||
Db        191 DILSGNIESDIDIDSLVTDKVVAALLSENEAGVNFARDIT---DIQGETHKADQDKIDIE 247

Qy         69 KTN----RESITTNVEGRRD--------IHKGHLEEKKDGSIKPEQKEDK 106
                |      :||  :|  ||          ||  :| : |   | ::: ||
Db        248 LDNIHESDSNITETIENLRDQLEKATDEEHKKEIESQVDAKKKQKEELDK 297


RESULT 20
US-10-770-726-49
; Sequence 49, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Brown, Eugene
;  APPLICANT:  Liu, Wei
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION:  CANCERS
;  FILE REFERENCE: AM101079 (031896-010000)
;  CURRENT APPLICATION NUMBER: US/10/770,726
;  CURRENT FILING DATE:  2004-02-04
;  NUMBER OF SEQ ID NOS: 48640
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
;   LENGTH: 795
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-770-726-49

  Query Match             11.3%;  Score 62.5;  DB 6;  Length 795;
  Best Local Similarity   25.9%;  Pred. No. 45;
  Matches   29;  Conservative   18;  Mismatches   46;  Indels   19;  Gaps    6;

Qy          2 KADTK--KNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSR 59
              ||: |  || : :    | | |    | |    :|:   |      |::     ::||
Db         29 KAEIKRLKNSDDRDSKRDSLEE---GELRDHRMEITIRNS-----PYRREDSMEDRGEED 80

Qy         60 DS------KEISIIEKTNRESITTNVEGRRDIHKGH-LEEKKDGSIKPEQKE 104
                ||      :::|  || :        | ||  |: |  |  |   :| :::|
Db         81 DSLAIKPPQQMSRKEKVHHRKDEKRKEKRR--HRSHSAEGGKHARVKEKERE 130


RESULT 21
US-11-196-475-70
; Sequence 70, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70
;   LENGTH: 663
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-70

  Query Match             11.2%;  Score 62;  DB 7;  Length 663;
  Best Local Similarity   27.2%;  Pred. No. 40;
  Matches   31;  Conservative   15;  Mismatches   50;  Indels   18;  Gaps    5;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEIS 65
              |||:       ||  : |     : |  | |||| |       :   | :| :  : :
Db        189 KKNILSGNVESDIDIDSLVTDKVVAA--LLSENESGVNFARDIT---DIQGETHKADQDK 243

Qy         66 I-IEKTN----RESITTNVEGRRD--------IHKGHLEEKKDGSIKPEQKEDK 106
              | ||  |      :||  :|  ||          ||  :| : |   | ::: ||
Db        244 IDIELDNFHESDSNITETIENLRDQLEKATDEEHKKEIESQVDAKKKQKEELDK 297


RESULT 22
US-11-196-475-78
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 78
;   LENGTH: 663
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-78

  Query Match             11.2%;  Score 62;  DB 7;  Length 663;
  Best Local Similarity   27.2%;  Pred. No. 40;
  Matches   31;  Conservative   15;  Mismatches   50;  Indels   18;  Gaps    5;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEIS 65
              |||:       ||  : |     : |  | |||| |       :   | :| :  : :
Db        189 KKNILSGNVESDIDIDSLVTDKVVAA--LLSENESGVNFARDIT---DIQGETHKADQDK 243

Qy         66 I-IEKTN----RESITTNVEGRRD--------IHKGHLEEKKDGSIKPEQKEDK 106
              | ||  |      :||  :|  ||          ||  :| : |   | ::: ||
Db        244 IDIELDNFHESDSNITETIENLRDQLEKATDEEHKKEIESQVDAKKKQKEELDK 297


RESULT 23
US-11-196-475-74
; Sequence 74, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
```

```
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-74

  Query Match             11.2%;  Score 62;  DB 7;  Length 700;
  Best Local Similarity   27.2%;  Pred. No. 43;
  Matches   31;  Conservative   15;  Mismatches   50;  Indels   18;  Gaps    5;

Qy          6 KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEIS 65
              |||:       ||  : |     : |  | |||| |       :   | :| :  : :
Db        189 KKNILSGNVESDIDIDSLVTDKVVAA--LLSENESGVNFARDIT---DIQGETHKADQDK 243

Qy         66 I-IEKTN----RESITTNVEGRRD--------IHKGHLEEKKDGSIKPEQKEDK 106
              | ||  |       :|| :|  ||         || :| : |   | ::: ||
Db        244 IDIELDNIHESDSNITETIENLRDQLEKATDEEHKKEIESQVDAKKKQKEELDK 297


RESULT 24
US-10-467-657-366
; Sequence 366, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 366
;   LENGTH: 451
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-366

  Query Match             11.1%;  Score 61.5;  DB 6;  Length 451;
  Best Local Similarity   25.0%;  Pred. No. 28;
  Matches   21;  Conservative   17;  Mismatches   33;  Indels   13;  Gaps    4;

Qy          8 NLERKKEHGDILAEDLYGRLEIPAIE----LPSENERGYYIPHQSSLPQDNRG----NSR 59
              | |: :|  :   | : |:   |  :    :   |  |  :|    : :|  |    ||:
Db        222 NFEKAQEEANQNFEGINGQKRHPYADKVFLIEESNRPGEMMP----IMEDEHGTYIMNSK 277

Qy         60 DSKEISIIEKTNRESI-TTNVEGR 82
              | : | ::||  :  : :  ||||
Db        278 DLRGIEVVEKLAKIGVDSLKVEGR 301
```

```
RESULT 25
US-10-821-234-1039
; Sequence 1039, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1039
;   LENGTH: 578
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1039

  Query Match             11.1%;  Score 61.5;  DB 6;  Length 578;
  Best Local Similarity   26.7%;  Pred. No. 38;
  Matches   24;  Conservative   16;  Mismatches   37;  Indels   13;  Gaps    4;

Qy         25 GRLEIPAIELPSENERGYYI--PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGR 82
              | |::     |: ::  | :   |||     :      : :|: | |    | |    : |
Db        184 GSLQLICKSEPNTDQLDYDVGEEHQSPGGISSEEEEEEEEEMLISE----EEIPFKDDPR 239

Qy         83 RDIHKGHLE------EKKDGSIKPEQKEDK 106
               : :| |||       :| | :| |:|| |
Db        240 DETYKPHLERETPKPRRKSGKVK-EEKEKK 268


RESULT 26
US-10-645-441-25
; Sequence 25, Application US/10645441
; Publication No. US20050260599A1
; GENERAL INFORMATION:
;  APPLICANT: Zuker, Charles S.
;  APPLICANT:  Ryba, Nicholas J.P.
;  APPLICANT:  Nelson, Greg
;  APPLICANT:  Hoon, Mark A.
;  APPLICANT:  Chandrashekar, Jayaram
;  APPLICANT:  Zhang, Yifeng
;  APPLICANT:  The Regents of the University of California
;  APPLICANT:  The Government of the United States of America
;  APPLICANT:  as represented by the Secretary of the
;  APPLICANT:  Department of Health and Human Services
;  TITLE OF INVENTION: Mammalian Sweet Taste Receptors
;  FILE REFERENCE: 02307E-120110US
;  CURRENT APPLICATION NUMBER: US/10/645,441
;  CURRENT FILING DATE:  2003-08-20
;  PRIOR APPLICATION NUMBER: US/09/927,315
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: US 60/302,898
;  PRIOR FILING DATE: 2001-07-03
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 858
;   TYPE: PRT
;   ORGANISM: Rattus sp.
;   FEATURE:
;   OTHER INFORMATION: rat T1R3 sweet taste receptor
US-10-645-441-25

  Query Match             11.1%;  Score 61.5;  DB 6;  Length 858;
  Best Local Similarity   26.8%;  Pred. No. 62;
  Matches   15;  Conservative   13;  Mismatches   27;  Indels    1;  Gaps    1;
```

```
Qy          20 AEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESI 75
               ::| ||| |  :|     | ||  | |:  :|| :    :  | : :: : |: :
Db         214 SDDDYGR-EGLSIFSGLANSRGICIAHEGLVPQHDTSGQQLGKVVDVLRQVNQSKV 268
```

RESULT 27
US-10-821-234-1120
; Sequence 1120, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1120
; LENGTH: 4384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1120

Query Match 11.1%; Score 61.5; DB 6; Length 4384;
Best Local Similarity 25.2%; Pred. No. 4.8e+02;
Matches 27; Conservative 18; Mismatches 47; Indels 15; Gaps 6;

```
Qy           2 KADTKKNLERKKE-HGDI--LAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNS 58
               |  |  :|   :|   |:  : |||      |   ::| |        | |  ||::  |
Db        1895 KLSTPSSLSSSQEILKDVAEMKEDLMRMTAILQTDVPEEK------PFQPELPKE--GRI 1946

Qy          59 RDSKEISIIEKTNRESITTNVEGRRDI---HKGHLEE-KKDGSIKPE 101
               | :   |:||   : :  :   ::|:   :||  :  | |    ||
Db        1947 DDEEPFKIVEKVKEDLVKVSEILKKDVCVDNKGSPKSPKSDKGHSPE 1993
```

RESULT 28
US-10-793-626-310
; Sequence 310, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 310
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-310

Query Match 11.0%; Score 61; DB 6; Length 250;
Best Local Similarity 28.6%; Pred. No. 15;
Matches 28; Conservative 13; Mismatches 23; Indels 34; Gaps 7;

```
Qy           5 TKKNLERKKEHGDILAEDLYGRL---EIPAIELPSENERGYYI---PHQSSLPQDNRGNS 58
               | | |:  ||  |::   || ||   ||  :   | : : :|    ||: ||||
Db          21 TLKGLKAIKE-ADVI---LYDRLVNKEI--LNYASPSTKFFYCGKDPHRHSLPQ------ 68

Qy          59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDG 96
                        |:||:  :|          |||:  :  |
Db          69 ---------EETNKMMVTL-------AKKGHIVTRLKG 90
```

RESULT 29
US-10-793-626-2174
; Sequence 2174, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2174
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2174

Query Match 11.0%; Score 61; DB 6; Length 250;
Best Local Similarity 28.6%; Pred. No. 15;
Matches 28; Conservative 13; Mismatches 23; Indels 34; Gaps 7;

```
Qy           5 TKKNLERKKEHGDILAEDLYGRL---EIPAIELPSENERGYYI---PHQSSLPQDNRGNS 58
               | | |:  ||  |::   || ||   ||  :   | : : :|    ||: ||||
Db          21 TLKGLKAIKE-ADVI---LYDRLVNKEI--LNYASPSTKFFYCGKDPHRHSLPQ------ 68

Qy          59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDG 96
                        |:||:  :|          |||:  :  |
Db          69 ---------EETNKMMVTL-------AKKGHIVTRLKG 90
```

RESULT 30
US-11-074-176-204
; Sequence 204, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 204
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-204

Query Match 11.0%; Score 61; DB 7; Length 552;
Best Local Similarity 22.9%; Pred. No. 40;
Matches 19; Conservative 16; Mismatches 34; Indels 14; Gaps 2;

```
Qy          12 KKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGN--SRDSKEISIIEK 69
               ||: | |:             :||  : :   :|| ::    | |    | :     |:
Db          49 KKQSGVIVT-----------DLPILHTKELGLPHDATQLVYNHGKILGRTASARHFIDN 96
```

```
Qy          70 TNRESITTNVEGRRDIHKGHLEE 92
                 | ::   :   | ||:||| ::
Db          97 PNEDAEALDANLREDIYKGHKQK 119
```

Search completed: December 15, 2005, 22:03:45
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 00:14:33 ; Search time 3943 Seconds
(without alignments)
1542.544 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103705_192/app_query.fasta_1.263
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_4939_@runat_15122005_103705_192 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	554	100.0	843	2	PFALSAC	L40885 Plasmodium
2	554	100.0	843	2	PFALSAH	L40890 Plasmodium
3	554	100.0	843	2	PFALSAI	L40891 Plasmodium
4	554	100.0	950	6	A28740	A28740 cDNA for LS
5	554	100.0	950	6	AR162961	AR162961 Sequence
6	554	100.0	950	6	AR178250	AR178250 Sequence
7	554	100.0	1482	6	AR162970	AR162970 Sequence
8	554	100.0	1482	6	AR178259	AR178259 Sequence
9	554	100.0	1493	6	A28743	A28743 cDNA for LS
10	554	100.0	1493	6	AR162967	AR162967 Sequence
11	554	100.0	1493	6	AR178256	AR178256 Sequence
12	554	100.0	1496	2	PFLSA13	Z30320 P.falciparu
13	553	99.8	843	2	PFALSAB	L40884 Plasmodium
14	553	99.8	843	2	PFALSAE	L40887 Plasmodium
15	553	99.8	843	2	PFALSAF	L40888 Plasmodium
16	553	99.8	843	2	PFALSAK	L40893 Plasmodium
17	553	99.8	1368	11	AY751501	AY751501 Synthetic
18	553	99.8	5970	2	PFLSA1G	X56203 P.falciparu
19	551	99.5	1482	6	AR162974	AR162974 Sequence
20	551	99.5	1482	6	AR178263	AR178263 Sequence
21	550	99.3	843	2	PFALSAG	L40889 Plasmodium
22	550	99.3	843	2	PFALSAL	L40834 Plasmodium
23	549	99.1	843	2	PFALSA1A	L40908 Plasmodium
24	549	99.1	843	2	PFALSA1B	L40909 Plasmodium
25	549	99.1	843	2	PFALSA1C	L40910 Plasmodium
26	549	99.1	843	2	PFALSA1N	L40947 Plasmodium
27	549	99.1	843	2	PFALSAD	L40886 Plasmodium
28	549	99.1	843	2	PFALSAJ	L40892 Plasmodium
29	549	99.1	843	2	PFALSAM	L40835 Plasmodium
30	549	99.1	843	2	PFALSAN	L40836 Plasmodium
31	549	99.1	843	2	PFALSAO	L40837 Plasmodium
32	548	98.9	253001	2	AE014834	AE014834 Plasmodiu
33	469	84.7	795	2	AF246996	AF246996 Plasmodiu
34	224	40.4	126	2	AY090552	AY090552 Plasmodiu
35	97	17.5	2130	6	AR549811	AR549811 Sequence
36	94.5	17.1	162642	5	BX908804	BX908804 Zebrafish
37	91	16.4	13097	1	AE000673	AE000673 Aquifex a
c 38	89.5	16.2	157332	14	AC069215	AC069215 Homo sapi
c 39	88.5	16.0	110000	1	BA000035_14	Continuation (15 o
40	88.5	16.0	208199	9	AC026231	AC026231 Mus muscu
c 41	88	15.9	217080	9	AC125565	AC125565 Rattus no
c 42	88	15.9	246668	14	AC103536	AC103536 Rattus no
c 43	88	15.9	249750	14	AC098941	AC098941 Rattus no
c 44	87.5	15.8	107112	8	AL596210	AL596210 Human DNA
45	87.5	15.8	184315	14	AC073482	AC073482 Homo sapi

ALIGNMENTS

RESULT 1
PFALSAC
LOCUS PFALSAC 843 bp DNA linear INV 19-AUG-1995
DEFINITION Plasmodium falciparum (clones KEN1-21b, KEN1-40h, KEN1-56b, PNG1-28a, PNG1-28c, BRA1-19e) liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION L40885
VERSION L40885.1 GI:725278
KEYWORDS exoerythrocytic; liver stage-specific antigen; liver stage-specific antigen 1; malaria.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 843)
AUTHORS Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M., Hawley,W.A., Collins,W.E. and Lal,A.A.
TITLE Sequence variations in the non-repetitive regions of the liver stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
PUBMED 7477115
COMMENT Original source text: Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN1-56b) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG1-28a) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG1-28c) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN1-21b) blood

stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN1-40h) blood stage parasite DNA; and Plasmodium falciparum (individual_isolate Kenyan) (clone: BRA1-19e) blood stage parasite DNA.
Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for Brazil. For each clone, the number following the KEN, PNG or BRA represents group number, the number following the '-' denotes the patient's blood sample number and the lower case letter represents recombinant clones from the corresponding isolates.

```
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN1-56b"
                     /clone="PNG1-28a"
                     /clone="PNG1-28c"
                     /clone="KEN1-21b"
                     /clone="KEN1-40h"
                     /clone="BRA1-19e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59231.1"
                     /db_xref="GI:950664"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       228
                     /gene="LSA-1"
                     /note="clone PNG1-28a"
                     /replace="t"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              3.88e-49        Length:         843
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAC (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

```
RESULT 2
PFALSAH
LOCUS       PFALSAH                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN6-21a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40890
VERSION     L40890.1  GI:725283
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN6-21a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN6-21a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41600.1"
                     /db_xref="GI:950669"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEYIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              3.88e-49        Length:         843
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAH (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120
```

```
Qy          41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy          61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy          81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy         101 GluGlnLysGluAspLysSer 107
               |||||||||||||||||||||
Db         301 GAACAAAAAGAAGATAAATCT 321
```

RESULT 3
PFALSAI

```
LOCUS       PFALSAI                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN7-40c) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40891
VERSION     L40891.1  GI:725284
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN7-40c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
            Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN7-40c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41601.1"
                     /db_xref="GI:950670"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDY
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:                3.88e-49        Length:          843
Score:                    554.00          Matches:         107
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-837-344-19 (1-107) x PFALSAI (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

RESULT 4
A28740

```
LOCUS       A28740                   950 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA-R-NR protein.
ACCESSION   A28740
VERSION     A28740.1  GI:1247512
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 29 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                4.41e-49        Length:          950
Score:                    554.00          Matches:         107
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-837-344-19 (1-107) x A28740 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 5
AR162961
LOCUS       AR162961                950 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 32 from patent US 6270771.
ACCESSION   AR162961
VERSION     AR162961.1  GI:16233425
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 32 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              4.41e-49       Length:         950
Score:                  554.00         Matches:        107
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-837-344-19 (1-107) x AR162961 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 6
AR178250
LOCUS       AR178250                950 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 32 from patent US 6319502.
ACCESSION   AR178250
VERSION     AR178250.1  GI:20219388
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 32 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              4.41e-49       Length:         950
Score:                  554.00         Matches:        107
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-837-344-19 (1-107) x AR178250 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 7
AR162970
LOCUS       AR162970                1482 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 41 from patent US 6270771.
ACCESSION   AR162970
VERSION     AR162970.1  GI:16233435
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 41 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              7.06e-49        Length:         1482
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-837-344-19 (1-107) x AR162970 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 8
AR178259
LOCUS       AR178259                1482 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 41 from patent US 6319502.
ACCESSION   AR178259
VERSION     AR178259.1  GI:20219397
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 41 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              7.06e-49        Length:         1482
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-837-344-19 (1-107) x AR178259 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 9
A28743
LOCUS       A28743                  1493 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA gene 3' (partial).
ACCESSION   A28743
VERSION     A28743.1  GI:1247516
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 32 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              7.12e-49        Length:         1493
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-837-344-19 (1-107) x A28743 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957
```

```
RESULT 10
AR162967
LOCUS       AR162967               1493 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 38 from patent US 6270771.
ACCESSION   AR162967
VERSION     AR162967.1  GI:16233432
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 38 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                 7.12e-49       Length:          1493
Score:                     554.00         Matches:         107
Percent Similarity:        100.00%        Conservative:    0
Best Local Similarity:     100.00%        Mismatches:      0
Query Match:               100.00%        Indels:          0
DB:                        6              Gaps:            0

US-09-837-344-19 (1-107) x AR162967 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 11
AR178256
LOCUS       AR178256               1493 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 38 from patent US 6319502.
ACCESSION   AR178256
VERSION     AR178256.1  GI:20219394
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 38 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                 7.12e-49       Length:          1493
Score:                     554.00         Matches:         107
Percent Similarity:        100.00%        Conservative:    0
Best Local Similarity:     100.00%        Mismatches:      0
Query Match:               100.00%        Indels:          0
DB:                        6              Gaps:            0

US-09-837-344-19 (1-107) x AR178256 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 12
PFLSA13
LOCUS       PFLSA13                1496 bp    DNA     linear   INV 05-SEP-1994
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen-1 (3').
ACCESSION   Z30320
VERSION     Z30320.1  GI:510183
KEYWORDS    liver stage antigen-1; LSA-1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     The Plasmodium falciparum liver stage antigen LSA-1 is well
            conserved and harbors major B- and T-cell epitopes
  JOURNAL   J. Immunol. (1994) In press
REFERENCE   2  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     Plasmodium falciparum liver stage antigen-1 is well conserved and
            contains potent B and T cell determinants
  JOURNAL   J. Immunol. 153 (1), 190-204 (1994)
   PUBMED   7515922
REFERENCE   3  (bases 1 to 1496)
  AUTHORS   Fidock,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1994) David A Fidock, Laboratory of bio-medical
            parasitology, Pasteur, Institute, 25, rue du Dr. Roux, PARIS, 75724
            CEDEX 15, France
FEATURES             Location/Qualifiers
     source          1. .1496
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="T9/96"
```

```
                     /db_xref="taxon:5833"
                     /clone="DG536, DG443, DG750"
                     /cell_type="human erythrocyte culture"
                     /clone_lib="Lambda gt11"
                     /dev_stage="asexual blood stage"
     gene            1. .1482
                     /gene="LSA-1"
     CDS             <1. .1482
                     /gene="LSA-1"
                     /note="nucl. 1 to 1496 corresponds to nucl. 4327 to 5822
                     of published NF54 strain
                     3' extremity of LSA-1 gene of strain T9/96"
                     /codon_start=1
                     /product="liver stage antigen-1"
                     /protein_id="CAA82975.1"
                     /db_xref="GI:510184"
                     /db_xref="UniProt/TrEMBL:Q25886"
                     /translation="QEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDL
                     EQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQRDLEQRKADT
                     KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKE
                     ISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKSADIQNHTLETVNIS
                     DVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIVQYDNFQDEENIGIYKELED
                     LIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDNNFKPNDKSLYDEHIKKYKN
                     DKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKYFMKL"
     repeat_region   1. .639
ORIGIN

Alignment Scores:
Pred. No.:              7.14e-49        Length:         1496
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFLSA13 (1-1496)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 13
PFALSAB
LOCUS       PFALSAB                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium flaciparum (clones KEN0-11a,-11b,-37d,-39a,-44e,-51a,
            -53c,-53d, -56a, PNG0-5h,-7a,-8f,-19a,-26b,-86h,-86i, BRA0-5a)
            liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40884
VERSION     L40884.1  GI:725277
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-44e) DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: KEN0-51a) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            PNG0-7a) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG0-5h) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            KEN0-11a) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: KEN0-11b) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            KEN0-37d) DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: KEN0-39a) DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-8f) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-19a) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-26b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-26b) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-86h) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-86i) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-53c) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: KEN0-53d) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-56a) blood stage parasite DNA; and Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA0-5a) blood stage
            parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES                Location/Qualifiers
     source             1. .843
                        /organism="Plasmodium falciparum"
                        /mol_type="genomic DNA"
                        /isolate="Kenyan"
                        /db_xref="taxon:5833"
                        /clone="KEN0-51a"
                        /clone="KEN0-44e"
                        /clone="PNG0-7a"
                        /clone="PNG0-5h"
                        /clone="KEN0-11a"
                        /clone="KEN0-11b"
                        /clone="KEN0-37d"
                        /clone="KEN0-39a"
                        /clone="PNG0-8f"
                        /clone="PNG0-19a"
                        /clone="PNG0-26b"
                        /clone="PNG0-86h"
                        /clone="PNG0-86i"
                        /clone="KEN0-53c"
                        /clone="KEN0-53d"
                        /clone="KEN0-56a"
                        /clone="BRA0-5a"
                        /dev_stage="blood stage parasite"
     gene               1. .843
                        /gene="LSA-1"
     CDS                <1. .843
                        /gene="LSA-1"
                        /codon_start=1
                        /product="liver stage-specific antigen 1"
```

```
                     /protein_id="AAB59230.1"
                     /db_xref="GI:950663"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       60
                     /gene="LSA-1"
                     /note="clone BRA0-5a"
                     /replace="g"
ORIGIN

Alignment Scores:
Pred. No.:              4.95e-49        Length:         843
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAB (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 14
PFALSAE
LOCUS       PFALSAE                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN3-39d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40887
VERSION     L40887.1  GI:725280
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN3-39d) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN3-39d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41597.1"
                     /db_xref="GI:950666"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDASLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:              4.95e-49        Length:         843
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAE (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 15
PFALSAF
LOCUS       PFALSAF                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN4-32a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40888
VERSION     L40888.1  GI:725281
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (clone: KEN4-32a) blood
            stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /clone="KEN4-32a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41598.1"
                     /db_xref="GI:950667"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKMKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:              4.95e-49        Length:         843
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAF (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

```
RESULT 16
PFALSAK
LOCUS       PFALSAK                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN9-37e) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40893
VERSION     L40893.1  GI:725286
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN9-37e) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN9-37e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41603.1"
                     /db_xref="GI:950672"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKDENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:              4.95e-49        Length:         843
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAK (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180
```

```
Qy          61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy          81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy         101 GluGlnLysGluAspLysSer 107
               |||||||||||||||||||||
Db         301 GAACAAAAAGAAGATAAATCT 321

RESULT 17
AY751501
LOCUS       AY751501               1368 bp    DNA     linear   SYN 29-MAR-2005
DEFINITION  Synthetic construct LSA-NRC and truncated LSA-NRC (LSA-1) gene,
            partial cds.
ACCESSION   AY751501
VERSION     AY751501.1  GI:58429869
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1368)
  AUTHORS   Hillier,C.J., Ware,L.A., Barbosa,A., Angov,E., Lyon,J.A.,
            Heppner,D.G. and Lanar,D.E.
  TITLE     Process Development and Analysis of Liver-Stage Antigen 1, a
            Preerythrocyte-Stage Protein-Based Vaccine for Plasmodium
            falciparum
  JOURNAL   Infect. Immun. 73 (4), 2109-2115 (2005)
   PUBMED   15784552
REFERENCE   2  (bases 1 to 1368)
  AUTHORS   Hillier,C.J., Ware,L.A., Barbosa,A., Angov,E., Lyon,J.A.,
            Heppner,D.G. and Lanar,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2004) Immunology, Walter Reed Army Institute of
            Research, 503 Robert Grant Ave, Silver Spring, MD 20910-7500, USA
FEATURES             Location/Qualifiers
     source          1. .1368
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     gene            1. .>1368
                     /gene="LSA-1"
     CDS             1. .>1368
                     /gene="LSA-1"
                     /note="liver stage antigen; recombinant protein expressed
                     in Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /product="LSA-NRC"
                     /protein_id="AAW78331.1"
                     /db_xref="GI:58429870"
                     /translation="MGTNSEKDEIIKSNLRSGSSNSRNRINEEKHEKKHVLSHNSYEK
                     TKNNENNKFFDKDKELTMSNVKNVSQTNFKSLLRNLGVSENIFLKENKLNKEGKLIEH
                     IINDDDDKKKYIKGQDENRQEDLEEKAAEQQSDLEQERLAKEKLQERLAKEKLQEQQR
                     DLEQRKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDN
                     RGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKSADIQN
                     HTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIVQYDNFQDEEN
                     IGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDNNFKPNDKSLY
                     DEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKYFMKLGGSGSP
                     HHHHHH"
     misc_feature    10. .1326
                     /gene="LSA-1"
                     /note="derived from Plasmodium falciparum 3D7"
     CDS             184. .>1368
                     /gene="LSA-1"
                     /codon_start=1
                     /transl_table=11
                     /product="truncated LSA-NRC"
                     /protein_id="AAW78332.1"
                     /db_xref="GI:58429871"
                     /translation="MSNVKNVSQTNFKSLLRNLGVSENIFLKENKLNKEGKLIEHIIN
                     DDDDKKKYIKGQDENRQEDLEEKAAEQQSDLEQERLAKEKLQERLAKEKLQEQQRDLE
                     QRKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGN
                     SRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKSADIQNHTL
                     ETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIVQYDNFQDEENIGI
                     YKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDNNFKPNDKSLYDEH
                     IKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKYFMKLGGSGSPHHH
                     HHH"
ORIGIN

Alignment Scores:
Pred. No.:              8.28e-49        Length:         1368
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     11              Gaps:           0

US-09-837-344-19 (1-107) x AY751501 (1-1368)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         493 CGCAAGGCTGACACGAAAAAAAACCTGGAACGCAAAAAGGAACACGGCGACGTTCTGGCT 552

Qy          21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         553 GAGGACCTGTACGGCCGCCTGGAAATCCCAGCTATCGAACTCCCATCCGAAAACGAACGC 612

Qy          41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         613 GGCTACTACATCCCACACCAGAGCAGCCTGCCACAAGATAATCGCGGGAACTCCCGCGAC 672

Qy          61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         673 AGTAAGGAAATCAGCATCATCGAAAAAACCAACCGCGAAAGCATTACCACCAACGTGGAA 732

Qy          81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         733 GGCCGCCGCGACATCCACAAAGGCCACCTCGAAGAAAAGAAAGACGGCTCCATCAAACCA 792

Qy         101 GluGlnLysGluAspLysSer 107
               |||||||||||||||||||||
Db         793 GAACAGAAAGAAGACAAAAGC 813

RESULT 18
PFLSA1G
LOCUS       PFLSA1G                5970 bp    DNA     linear   INV 18-APR-2005
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen.
ACCESSION   X56203 S75010 S75012 S75014 S75016
VERSION     X56203.1  GI:9915
KEYWORDS    liver stage antigen.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Zhu,J. and Hollingdale,M.R.
  TITLE     Structure of Plasmodium falciparum liver stage antigen-1
  JOURNAL   Mol. Biochem. Parasitol. 48 (2), 223-226 (1991)
   PUBMED   1840628
REFERENCE   2  (bases 1 to 5970)
  AUTHORS   Zhu,J. and Hollingdale,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1990) Zhu J., Hollingdale M, Biomedical Research
            Institute, 12111 Parklawn Drive, Rockeville, MD 20852, U.S.A
COMMENT     On or before Jun 27, 2005 this sequence version replaced gi:241572,
            gi:241573, gi:241574, gi:241575.
FEATURES             Location/Qualifiers
     source          1. .5970
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="NF54"
                     /db_xref="taxon:5833"
                     /clone="RI-LSA segment, DRAI-LSA segment"
```

```
                     /cell_line="NF54 blood stage culture"
                     /clone_lib="pUC13 plasmid, RI-sublib, DRAI-sublib"
                     /dev_stage="liver stage"
     TATA_signal     30. .36
     gene            79. .5808
                     /gene="LSA-1"
     CDS             79. .5808
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage antigen"
                     /protein_id="CAA39663.1"
                     /db_xref="GI:9916"
                     /db_xref="HSSP:1D7M"
                     /db_xref="UniProt/TrEMBL:Q25893"
                     /translation="MKHILYISFYFILVNLLIFHINGKIIKNSEKDEIIKSNLRSGSS
                     NSRNRINEEKHEKKHVLSHNSYEKTKNNENNKFFDKDKELTMSNVKNVSQTNFKSLLR
                     NLGVSENIFLKENKLNKEGKLIEHIINDDDDKKKYIKGQDENRQEDLEEKAAKETLQG
                     QQSDLEQERLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQEQQSDLEQDRLA
                     KEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKL
                     QEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDL
                     EQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLERTKASKETLQEQQSDLEQERLAK
                     EKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQ
                     SDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDR
                     LAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQ
                     EQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLE
                     QERLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKE
                     KLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQS
                     DLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERL
                     AKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQE
                     QQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLERTKASKETLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRA
                     KEKLQEQQSDLEQDRLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDLEQE
                     RRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLANEKLQEQQRDLEQERRAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDL
                     EQERLAKEKLQEQQRDLEQRKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENE
                     RGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGS
                     IKPEQKEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDE
                     FKPIVQYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYE
                     KTKDNNFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSE
                     DITKYFMKL"
ORIGIN

Alignment Scores:
Pred. No.:              3.94e-48        Length:         5970
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFLSA1G (5970)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       4966 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 5025

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5026 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 5085

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5086 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 5145

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       5146 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 5205

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5206 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 5265

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db       5266 GAACAAAAAGAAGATAAATCT 5286

RESULT 19
AR162974
LOCUS       AR162974               1482 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 45 from patent US 6270771.
ACCESSION   AR162974
VERSION     AR162974.1  GI:16233440
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 45 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.46e-48        Length:         1482
Score:                  551.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.46%          Indels:         0
DB:                     6               Gaps:           0

US-09-837-344-19 (1-107) x AR162974 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 20
AR178263
LOCUS       AR178263               1482 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 45 from patent US 6319502.
ACCESSION   AR178263
VERSION     AR178263.1  GI:20219401
KEYWORDS    .
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 45 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.46e-48        Length:         1482
Score:                  551.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.46%          Indels:         0
DB:                     6               Gaps:           0

US-09-837-344-19 (1-107) x AR178263 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 21
PFALSAG
LOCUS       PFALSAG                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN5-44d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40889
VERSION     L40889.1  GI:725282
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN5-44d) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
            Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN5-44d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41599.1"
                     /db_xref="GI:950668"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPKQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:              1.03e-48        Length:         843
Score:                  550.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.28%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAG (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              :::|||||||||||||||||||
Db        301 AAACAAAAAGAAGATAAATCT 321

RESULT 22
PFALSAL
LOCUS       PFALSAL                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone PNG3-5f,-68f,-68h) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40834
VERSION     L40834.1  GI:725287
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-5f) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG3-68h) blood
            stage parasite DNA; and Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-68f) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="PNG3-5f"
                     /clone="PNG3-68h"
                     /clone="PNG3-68f"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59233.1"
                     /db_xref="GI:950673"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEE
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       573
                     /gene="LSA-1"
                     /note="clone PNG3-68h"
                     /replace="c"
ORIGIN

Alignment Scores:
Pred. No.:              1.03e-48        Length:         843
Score:                  550.00          Matches:        105
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  98.13%          Mismatches:     0
Query Match:            99.28%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSAL (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||:::|||||||||||||||
Db        301 GAAGAAAAAGAAGATAAATCT 321

RESULT 23
PFALSA1A
LOCUS       PFALSA1A                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40908
VERSION     L40908.1  GI:725293
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA7-30a) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA7-30a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59228.1"
                     /db_xref="GI:725294"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFRPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       240
                     /gene="LSA-1"
                     /note="clone BRA7-30a"
                     /replace="g"
ORIGIN

Alignment Scores:
Pred. No.:              1.31e-48        Length:         843
Score:                  549.00          Matches:        105
Percent Similarity:     99.07%          Conservative:   1
Best Local Similarity:  98.13%          Mismatches:     1
Query Match:            99.10%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x PFALSA1A (1-843)
```

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAG 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

RESULT 24
PFALSA1B
LOCUS PFALSA1B 843 bp DNA linear INV 19-AUG-1995
DEFINITION Plasmodium falciparum (clone BRA8-30f) liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION L40909
VERSION L40909.1 GI:725295
KEYWORDS exoerythrocytic; liver stage-specific antigen; liver stage-specific antigen 1; malaria.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 843)
AUTHORS Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M., Hawley,W.A., Collins,W.E. and Lal,A.A.
TITLE Sequence variations in the non-repetitive regions of the liver stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
PUBMED 7477115
COMMENT Original source text: Plasmodium falciparum (individual_isolate Brazilian) (clone: BRA8-30f) blood stage parasite DNA.
Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for Brazil. For each clone, the number following the KEN, PNG or BRA represents group number, the number following the '-' denotes the patient's blood sample number and the lower case letter represents recombinant clones from the corresponding isolates.
FEATURES Location/Qualifiers
source 1. .843
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/isolate="Brazilian"
/db_xref="taxon:5833"
/clone="BRA8-30f"
/dev_stage="blood stage parasite"
gene 1. .843
/gene="LSA-1"
CDS <1. .843
/gene="LSA-1"
/codon_start=1
/product="liver stage-specific antigen 1"
/protein_id="AAB59229.1"
/db_xref="GI:725296"
/translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVGELSEDITKY
FMKL"
variation 240
/gene="LSA-1"
/note="clone BRA8-30f"
/replace="a"
ORIGIN

Alignment Scores:

Pred. No.:	1.31e-48	Length:	843
Score:	549.00	Matches:	105
Percent Similarity:	99.07%	Conservative:	1
Best Local Similarity:	98.13%	Mismatches:	1
Query Match:	99.10%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-19 (1-107) x PFALSA1B (1-843)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAG 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

RESULT 25
PFALSA1C
LOCUS PFALSA1C 843 bp DNA linear INV 19-AUG-1995
DEFINITION Plasmodium falciparum (clone BRA9-32b) liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION L40910
VERSION L40910.1 GI:725297
KEYWORDS exoerythrocytic; liver stage-specific antigen; liver stage-specific antigen 1; malaria.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 843)
AUTHORS Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M., Hawley,W.A., Collins,W.E. and Lal,A.A.
TITLE Sequence variations in the non-repetitive regions of the liver stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
PUBMED 7477115
COMMENT Original source text: Plasmodium falciparum (individual_isolate Brazilian) (clone: BRA9-32b) blood stage parasite DNA.
Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for Brazil. For each clone, the number following the KEN, PNG or BRA represents group number, the number following the '-' denotes the patient's blood sample number and the lower case letter represents recombinant clones from the corresponding isolates.
FEATURES Location/Qualifiers
source 1. .843
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/isolate="Brazilian"

```
                     /db_xref="taxon:5833"
                     /clone="BRA9-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41582.1"
                     /db_xref="GI:725298"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e-48       Length:          843
Score:                    549.00         Matches:         105
Percent Similarity:       99.07%         Conservative:    1
Best Local Similarity:    98.13%         Mismatches:      1
Query Match:              99.10%         Indels:          0
DB:                       2              Gaps:            0

US-09-837-344-19 (1-107) x PFALSA1C (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 26
PFALSA1N
LOCUS       PFALSA1N                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40947
VERSION     L40947.1  GI:725317
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115

COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA6-29c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA6-29c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41592.1"
                     /db_xref="GI:725318"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVSDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e-48       Length:          843
Score:                    549.00         Matches:         105
Percent Similarity:       99.07%         Conservative:    1
Best Local Similarity:    98.13%         Mismatches:      1
Query Match:              99.10%         Indels:          0
DB:                       2              Gaps:            0

US-09-837-344-19 (1-107) x PFALSA1N (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 27
PFALSAD
LOCUS       PFALSAD                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN2-51b, PNG2-7f,-9f,-11a,-11b,-19d,
            -26e, BRA2-14c,-30b,-30) liver stage-specific antigen 1 (LSA-1)
            gene, 3' end of cds.
ACCESSION   L40886
VERSION     L40886.1  GI:725279
```

```
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN2-51b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-30e) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG2-26e) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-14c) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA2-30b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG2-7f) blood stage
            parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: PNG2-9f) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-11a) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            PNG2-11b) blood stage parasite DNA; and Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-19d) blood stage parasite
            DNA.
            1. Last author is Altaf A. Lal 2. Following clones also have the sa
            me nucleotide sequence: KEN121b, KEN140h, KEN156b, PNG1281 (at 5193
            A to T), P
            NG128c, BRA119e
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN2-51b"
                     /clone="BRA2-30e"
                     /clone="PNG2-26e"
                     /clone="BRA2-14c"
                     /clone="BRA2-30b"
                     /clone="PNG2-7f"
                     /clone="PNG2-9f"
                     /clone="PNG2-11a"
                     /clone="PNG2-11b"
                     /clone="PNG2-19d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59232.1"
                     /db_xref="GI:950665"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       222
                     /gene="LSA-1"
                     /note="clone PNG2-9f"
                     /replace="c"
     variation       240
                     /gene="LSA-1"
                     /note="clones PNG2-7f, BRA2-30b and BRA2-30e"
                     /replace="g"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.31e-48	Length:	843
Score:	549.00	Matches:	105
Percent Similarity:	99.07%	Conservative:	1
Best Local Similarity:	98.13%	Mismatches:	1
Query Match:	99.10%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-19 (1-107) x PFALSAD (1-843)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321
```

```
RESULT 28
PFALSAJ
LOCUS       PFALSAJ                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN8-32b) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40892
VERSION     L40892.1  GI:725285
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN8-32b) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
```

```
                     /db_xref="taxon:5833"
                     /clone="KEN8-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41602.1"
                     /db_xref="GI:950671"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e-48       Length:          843
Score:                    549.00         Matches:         105
Percent Similarity:       99.07%         Conservative:    1
Best Local Similarity:    98.13%         Mismatches:      1
Query Match:              99.10%         Indels:          0
DB:                       2              Gaps:            0

US-09-837-344-19 (1-107) x PFALSAJ (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 29
PFALSAM
LOCUS       PFALSAM                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clones BRA3-5b,-19a,-14b,-28e) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40835
VERSION     L40835.1  GI:725288
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115

COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA3-14b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA3-5b) blood stage
            parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: BRA3-28e) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: BRA3-29b) blood stage parasite
            DNA; and Plasmodium falciparum (individual_isolate Kenyan) (clone:
            BRA3-19a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal 2. The following clones have the sam
            e nucleotide sequence: BRA3-5b; BRA3-19a; BRA3-14b (one silent
            change at 5181,
            A to G) and BRA3-28e (one silent change at 5040, T to C) Clone name
            abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for
            Brazil. For each clone, the number following the KEN, PNG or BRA
            represents group number, the number following the '-' denotes the
            patient's blood sample number and the lower case letter represents
            recombinant clones from the corresponding isolates.
FEATURES              Location/Qualifiers
     source           1. .843
                      /organism="Plasmodium falciparum"
                      /mol_type="genomic DNA"
                      /isolate="Kenyan"
                      /db_xref="taxon:5833"
                      /clone="BRA3-14b"
                      /clone="BRA3-5b"
                      /clone="BRA3-28e"
                      /clone="BRA3-29b"
                      /clone="BRA3-19a"
                      /dev_stage="blood stage parasite"
     gene             1. .843
                      /gene="LSA-1"
     CDS              <1. .843
                      /gene="LSA-1"
                      /codon_start=1
                      /product="liver stage-specific antigen 1"
                      /protein_id="AAB59234.1"
                      /db_xref="GI:950674"
                      /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                      PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                      KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                      QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                      NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                      FMKL"
     variation        75
                      /gene="LSA-1"
                      /note="clone BRA3-28e"
                      /replace="c"
     variation        216
                      /gene="LSA-1"
                      /note="clone BRA3-14b"
                      /replace="g"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e-48       Length:          843
Score:                    549.00         Matches:         105
Percent Similarity:       99.07%         Conservative:    1
Best Local Similarity:    98.13%         Mismatches:      1
Query Match:              99.10%         Indels:          0
DB:                       2              Gaps:            0

US-09-837-344-19 (1-107) x PFALSAM (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321

RESULT 30
PFALSAN
LOCUS       PFALSAN                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA4-28a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40836
VERSION     L40836.1  GI:725289
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA4-28a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="BRA4-28a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41604.1"
                     /db_xref="GI:950675"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEERIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

Alignment Scores:
Pred. No.:              1.31e-48        Length:         843
Score:                  549.00          Matches:        105
Percent Similarity:     99.07%          Conservative:   1
Best Local Similarity:  98.13%          Mismatches:     1
Query Match:            99.10%          Indels:         0
DB:                     2               Gaps:           0
```

```
US-09-837-344-19 (1-107) x PFALSAN (1-843)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||    ||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCT 321


Search completed: December 20, 2005, 02:40:25
Job time : 3954 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 00:10:53 ; Search time 495 Seconds
(without alignments)
1440.651 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103704_182/app_query.fasta_1.263
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_727_@runat_15122005_103704_182 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	950	2	AAQ28115	Aaq28115 P.falcipa
2	554	100.0	1496	2	AAQ28119	Aaq28119 P.falcipa
3	553	99.8	1320	2	AAQ80916	Aaq80916 Plasmodiu
4	553	99.8	1371	12	ADO21941	Ado21941 LSA-NRC(H
5	553	99.8	1374	12	ADO21919	Ado21919 LSA-NRC(H
6	94	17.0	1384	14	ADZ40618	Adz40618 PfCTL/HTL
7	94	17.0	1385	9	ADA49438	Ada49438 Multi-epi
8	94	17.0	1385	12	ADO24116	Ado24116 Epigene c
9	82	14.8	2930	14	AEA26642	Aea26642 Stress to
10	81.5	14.7	696	4	AAH84231	Aah84231 Human cel
11	81.5	14.7	699	4	AAH84230	Aah84230 Human cel
12	81.5	14.7	717	4	AAH84229	Aah84229 Human cel
13	81.5	14.7	774	4	AAH84228	Aah84228 Human cel
14	81.5	14.7	819	4	AAH84227	Aah84227 Human cel
15	81.5	14.7	1137	6	ABN66837	Abn66837 Streptoco
16	81.5	14.7	1239	13	ADU69448	Adu69448 S agalact
17	81.5	14.7	1239	13	ADV85196	Adv85196 Streptoco
18	81.5	14.7	1455	3	AAA05838	Aaa05838 Group B S
19	81.5	14.7	1669	4	AAH84226	Aah84226 Human cel
20	81.5	14.7	2460	5	AAS74470	Aas74470 DNA encod
c 21	81.5	14.7	75248	13	ADV87739	Adv87739 Streptoco
c 22	81.5	14.7	75248	13	ADV78992	Adv78992 Streptoco
c 23	81.5	14.7	110000	6	ABN71527_18	Continuation (19 o
c 24	81.5	14.7	110000	13	ADV81204_19	Continuation (20 o
c 25	81.5	14.7	110000	13	ADV81204_20	Continuation (21 o
26	81	14.6	1508	13	ADX11745	Adx11745 Plant ful
27	80.5	14.5	540	5	ABV49476	Abv49476 Human pro
28	80.5	14.5	1116	14	ADZ72254	Adz72254 Plasmodiu
29	80.5	14.5	1353	2	AAV74832	Aav74832 Staphyloc
30	80.5	14.5	1407	8	ACF73779	Acf73779 Staphyloc
31	80.5	14.5	1524	6	AAD33372	Aad33372 Staphyloc
32	80.5	14.5	2709	8	ACA28274	Aca28274 Prokaryot
33	80	14.4	1685	10	ADB63803	Adb63803 Human cDN
34	80	14.4	2193	13	ADT16091	Adt16091 Plant cDN
c 35	80	14.4	3684	4	ABL23583	Abl23583 Drosophil
36	79.5	14.4	700	2	AAV88907	Aav88907 EST clone
37	79.5	14.4	966	4	AAI19757	Aai19757 Probe #96
38	79.5	14.4	966	4	ABA64784	Aba64784 Human foe
39	79.5	14.4	966	4	AAI44953	Aai44953 Probe #13
40	79.5	14.4	966	4	ABA46903	Aba46903 Human bre
41	79.5	14.4	966	4	ABA31902	Aba31902 Probe #10
42	79.5	14.4	966	4	AAK38947	Aak38947 Human bon
43	79.5	14.4	966	4	AAK13216	Aak13216 Human bra
44	79.5	14.4	966	4	ABS38535	Abs38535 Human liv
45	79.5	14.4	966	5	AAI05474	Aai05474 Probe #54

ALIGNMENTS

RESULT 1
AAQ28115
ID AAQ28115 standard; DNA; 950 BP.
XX
AC AAQ28115;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA-R-NR coding sequence.
XX
KW Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW paludism; liver stage-specific antigen; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT repeat_region 3. .629
FT /*tag= a
FT /rpt_type= TANDEM
FT repeat_unit 3. .53
FT /*tag= b
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.

```
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   P-PSDB; AAR26941.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 17; Fig 2; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert (950 bases) encoded LSA-R-NR
CC   containing a 12-repeat region followed by a non-repeat region. Preferred
CC   antigenic polypeptides of the invention are derived from the amino acid
CC   sequence of LSA-R-NR. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 950 BP; 456 A; 127 C; 206 G; 161 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.47e-59      Length:         950
Score:                  554.00        Matches:        107
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-837-344-19 (1-107) x AAQ28115 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 2
AAQ28119
ID   AAQ28119 standard; DNA; 1496 BP.
XX
AC   AAQ28119;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA gene 3' region.
XX
KW   Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW   paludism; liver stage-specific antigen; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1482
FT                   /*tag=  c
FT                   /product= "LSA_C-terminal_region"
FT   repeat_region   37. .639
FT                   /*tag=  a
FT   repeat_unit     37. .87
FT                   /*tag=  b
FT                   /rpt_type= TANDEM
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   P-PSDB; AAR26944.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 2; Fig 8-10; 81pp; French.
XX
CC   This nucleotide sequence is the 3' part of the P.falciparum liver-stage
CC   specific antigen (LSA) gene. It codes for a polypeptide sequence which
CC   carries a T cell epitope characteristic of a protein produced in
CC   hepatocytes infected with P.falciparum. The polypeptide can be used in
CC   the preparation of vaccines against malaria. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 1496 BP; 716 A; 169 C; 300 G; 311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.48e-59      Length:         1496
Score:                  554.00        Matches:        107
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-837-344-19 (1-107) x AAQ28119 (1-1496)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
```

```
Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 3
AAQ80916
ID   AAQ80916 standard; cDNA; 1320 BP.
XX
AC   AAQ80916;
XX
DT   16-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   29-AUG-1995  (first entry)
XX
DE   Plasmodium falciparum LSA-1 repeatless gene cDNA.
XX
KW   Plasmodium falciparum LSA-1 repeatless gene; recombinant poxvirus;
KW   multicomponent multistage malarial vaccines; immunogens;
KW   malaria diagnosis; ss.
XX
OS   Plasmodium falciparum; (pLSARPLS.I4L.1).
XX
PN   WO9428930-A1.
XX
PD   22-DEC-1994.
XX
PF   10-JUN-1994;   94WO-US006652.
XX
PR   11-JUN-1993;   93US-00075783.
PR   09-JUN-1994;   94US-00257073.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Paoletti E,  De Taisne C,  Tine JA;
XX
DR   WPI; 1995-036113/05.
XX
PT   Recombinant poxvirus contg. Plasmodium DNA in non-essential region -
PT   useful in vaccines against malaria and for prodn. of Plasmodium
PT   immunogens.
XX
PS   Claim 3; Fig 11; 183pp; English.
XX
CC   AAQ80916 is the P. falciparum LSA-1 repeatless gene cDNA sequence. New
CC   recombinant poxviruses containing either the SERA, ABRA, Pfhsp70, AMA-1,
CC   Pfs25, Pfs16, CSP, PfSSP2, LSA-1, LSA-1 repeatless, MSA-1, MSA-1 (N-
CC   terminal p83 or C-terminal gp42) genes, or a combination of these in non-
CC   essential regions of their genomes are claimed. These poxviruses (pref.
CC   with a virulence reducing genomic deletion or disruption) can be used as
CC   vaccines against malaria and for the prodn. of Plasmodium immunogens.
CC   These viruses provide multicomponent, multistage vaccines due to their
CC   expression of sporozite, liver stage, blood stage and sexual stage
CC   proteins. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 16-
CC   OCT-2003 to standardise OS field)
XX
SQ   Sequence 1320 BP; 619 A; 128 C; 227 G; 346 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.07e-59        Length:         1320
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x AAQ80916 (1-1320)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        478 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 537

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 597

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 657

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 717

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 777

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||
Db        778 GAACAAAAAGAAGATAAATCT 798

RESULT 4
ADO21941
ID   ADO21941 standard; DNA; 1371 BP.
XX
AC   ADO21941;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H) construct DNA derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H) construct; ds; gene.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1371
FT                   /*tag= a
FT                   /product= "LSA-NRC(H) construct protein derived from
FT                   Malaria parasite LSA-1"
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   P-PSDB; ADO21942.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 16; SEQ ID NO 25; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct DNA of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
```

```
CC   parasite P. falciparum LSA-1 in addition to a C-terminal His6 tag.
XX
SQ   Sequence 1371 BP; 539 A; 358 C; 283 G; 191 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.33e-59       Length:        1371
Score:                  553.00         Matches:       106
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  99.07%         Mismatches:    0
Query Match:            99.82%         Indels:        0
DB:                     12             Gaps:          0

US-09-837-344-19 (1-107) x ADO21941 (1-1371)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        493 CGCAAGGCTGACACGAAAAAAAACCTGGAACGCAAAAAGGAACACGGCGACGTTCTGGCT 552

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 GAGGACCTGTACGGCCGCCTGGAAATCCCAGCTATCGAACTCCCATCCGAAAACGAACGC 612

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 GGCTACTACATCCCACACCAGAGCAGCCTGCCACAAGATAATCGCGGGAACTCCCGCGAC 672

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        673 AGTAAGGAAATCAGCATCATCGAAAAAACCAACCGCGAAAGCATTACCACCAACGTGGAA 732

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 GGCCGCCGCGACATCCACAAAGGCCACCTCGAAGAAAAGAAAGACGGCTCCATCAAACCA 792

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        793 GAACAGAAAGAAGACAAAAGC 813

RESULT 5
ADO21919
ID   ADO21919 standard; DNA; 1374 BP.
XX
AC   ADO21919;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H)Mut construct DNA derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H)Mut construct; ds; gene;
KW   mutant.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1374
FT                   /*tag=  a
FT                   /product= "LSA-NRC(H)Mut construct protein derived from
FT                   Malaria parasite LSA-1"
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   P-PSDB; ADO21920.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 17; SEQ ID NO 3; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct DNA of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to an amino acid insertion
CC   within the T5 P. falciparum epitope and a C-terminal His6 tag.
XX
SQ   Sequence 1374 BP; 539 A; 358 C; 285 G; 192 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.34e-59       Length:        1374
Score:                  553.00         Matches:       106
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  99.07%         Mismatches:    0
Query Match:            99.82%         Indels:        0
DB:                     12             Gaps:          0

US-09-837-344-19 (1-107) x ADO21919 (1-1374)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        493 CGCAAGGCTGACACGAAAAAAAACCTGGAACGCAAAAAGGAACACGGCGACGTTCTGGCT 552

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 GAGGACCTGTACGGCCGCCTGGAAATCCCAGCTATCGAACTCCCATCCGAAAACGAACGC 612

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 GGCTACTACATCCCACACCAGAGCAGCCTGCCACAAGATAATCGCGGGAACTCCCGCGAC 672

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        673 AGTAAGGAAATCAGCATCATCGAAAAAACCAACCGCGAAAGCATTACCACCAACGTGGAA 732

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 GGCCGCCGCGACATCCACAAAGGCCACCTCGAAGAAAAGAAAGACGGCTCCATCAAACCA 792

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        793 GAACAGAAAGAAGACAAAAGC 813

RESULT 6
ADZ40618
ID   ADZ40618 standard; DNA; 1384 BP.
XX
AC   ADZ40618;
XX
DT   30-JUN-2005  (first entry)
XX
DE   PfCTL/HTL/N multi-epitope construct encoding DNA.
XX
KW   multi-epitope; immune stimulation; HIV infection; anti-HIV;
KW   immunostimulant; vaccine; immunogenicity; gene; ds.
XX
OS   Synthetic.
```

XX
PN WO2005033265-A2.
XX
PD 14-APR-2005.
XX
PF 26-APR-2004; 2004WO-US012732.
XX
PR 25-APR-2003; 2003US-0465229P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Chesnut RW, Newman MJ, Livingston BD;
XX
DR WPI; 2005-285416/29.
DR P-PSDB; ADZ40617.
XX
PT New multi-epitope polynucleotide, useful for designing vaccines that
PT provide increased immunogenicity, and for inducing immune responses
PT against HIV in an individual.
XX
PS Disclosure; Fig 18I; 261pp; English.
XX
CC The invention relates to a novel multi-epitope polynucleotide. The
CC invention further comprises: a composition comprising the polynucleotide;
CC a cell comprising the polynucleotide; a method for inducing an immune
CC response against human immunodeficiency virus (HIV) in an individual,
CC comprising administering the polynucleotide, composition or cell to the
CC individual; and a method for making the polynucleotide, composition or
CC cell. The multi-epitope polynucleotide has anti-HIV and immunostimulant
CC activities. The polynucleotide is useful for designing vaccines that
CC provide increased immunogenicity, and for inducing an immune response
CC against HIV. This polynucleotide sequence represents an PfCTL/HTL/N multi
CC -epitope construct encoding DNA of the invention.
XX
SQ Sequence 1384 BP; 351 A; 347 C; 361 G; 325 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.123	Length:	1384
Score:	94.00	Matches:	18
Percent Similarity:	85.71%	Conservative:	0
Best Local Similarity:	85.71%	Mismatches:	3
Query Match:	16.97%	Indels:	0
DB:	14	Gaps:	0

US-09-837-344-19 (1-107) x ADZ40618 (1-1384)

```
Qy         30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
              |||         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 CCTGGACCAGGTCTGCCGTCCGAAAACGAAAGAGGATACTACATACCTCACCAGAGCAGC 419

Qy         50 Leu 50
              |||
Db        420 CTC 422
```

RESULT 7
ADA49438
ID ADA49438 standard; DNA; 1385 BP.
XX
AC ADA49438;
XX
DT 20-NOV-2003 (first entry)
XX
DE Multi-epitope construct #1 DNA.
XX
KW ds; gene; multi-epitope; immunogenic; epitope;
KW major histocompatibility complex; MHC class I; MHC class II;
KW junctional epitope.
XX
OS Synthetic.
OS Plasmodium falciparum.
XX
PN US2002119127-A1.
XX
PD 29-AUG-2002.
XX
PF 27-JUN-2001; 2001US-00894018.
XX
PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.
XX
PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.
XX
PI Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;
XX
DR WPI; 2003-615704/58.
DR P-PSDB; ADA49437.
XX
PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.
XX
PS Disclosure; Fig 18I; 78pp; English.
XX
CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer
CC residues between two CEN, and selecting the constructs having less
CC junctional epitopes. The method is useful for designing a multi-epitope
CC construct having multiple epitope nucleic acid. The method avoids or
CC minimises the occurrence of junctional epitopes and maximises the
CC immunogenicity and/or antigenicity of multi-epitope vaccines. The present
CC sequence represents DNA encoding a multi-epitope construct.
XX
SQ Sequence 1385 BP; 351 A; 347 C; 361 G; 326 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.124	Length:	1385
Score:	94.00	Matches:	18
Percent Similarity:	85.71%	Conservative:	0
Best Local Similarity:	85.71%	Mismatches:	3
Query Match:	16.97%	Indels:	0
DB:	9	Gaps:	0

US-09-837-344-19 (1-107) x ADA49438 (1-1385)

```
Qy         30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
              |||         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTGGACCAGGTCTGCCGTCCGAAAACGAAAGAGGATACTACATACCTCACCAGAGCAGC 420

Qy         50 Leu 50
              |||
Db        421 CTC 423
```

RESULT 8
ADO24116
ID ADO24116 standard; DNA; 1385 BP.
XX
AC ADO24116;
XX
DT 01-JUL-2004 (first entry)
XX
DE Epigene construct PfCTL/HTL(N) DNA.
XX
KW ds; gene; hepatotropic; virucide; antiinflammatory; anti-HIV; cytostatic;
KW epitope; cytotoxic T lymphocyte; CTL; helper T lymphocyte; HTL;
KW immune response; hepatitis B virus; HBV; HIV; cancer; lymphoma.
XX
OS Unidentified.

```
XX
FH   Key             Location/Qualifiers
FT   exon            1. .1377
FT                   /*tag=  a
XX
PN   WO2004031210-A2.
XX
PD   15-APR-2004.
XX
PF   03-OCT-2003; 2003WO-US031303.
XX
PR   03-OCT-2002; 2002US-0415463P.
PR   22-OCT-2002; 2002US-0419973P.
XX
PA   (EPIM-) EPIMMUNE INC.
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Sette A,  Chesnut R,  Newman MJ,  Livingston BD,  Babe LM,  Chen Y;
PI   Deyoung LM,  Huang MTF,  Power SD;
XX
DR   WPI; 2004-330143/30.
DR   P-PSDB; ADO24115.
XX
PT   New multi-epitope polynucleotides encoding cytotoxic T lymphocyte and/or
PT   helper T lymphocyte epitopes, useful for inducing or stimulating an
PT   antiviral or anticancer immune response, especially against hepatitis B
PT   virus.
XX
PS   Disclosure; SEQ ID NO 266; 401pp; English.
XX
CC   The invention relates to a multi-epitope polynucleotide (I) encoding
CC   cytotoxic T lymphocyte (CTL) and/or helper T lymphocyte (HTL) epitopes.
CC   The composition and methods are useful for inducing or stimulating an
CC   immune response against hepatitis B virus (HBV), other viruses (e.g.
CC   HIV), or against cancers such as prostate cancer or lymphoma. This
CC   sequence corresponds to an example of an epigene construct used in the
CC   invention.
XX
SQ   Sequence 1385 BP; 351 A; 347 C; 361 G; 326 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.124          Length:        1385
Score:                  94.00          Matches:       18
Percent Similarity:     85.71%         Conservative:  0
Best Local Similarity:  85.71%         Mismatches:    3
Query Match:            16.97%         Indels:        0
DB:                     12             Gaps:          0

US-09-837-344-19 (1-107) x ADO24116 (1-1385)

Qy         30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
              |||          |||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTGGACCAGGTCTGCCGTCCGAAAACGAAAGAGGATACTACATACCTCACCAGAGCAGC 420

Qy         50 Leu 50
              |||
Db        421 CTC 423

RESULT 9
AEA26642
ID   AEA26642 standard; DNA; 2930 BP.
XX
AC   AEA26642;
XX
DT   28-JUL-2005  (first entry)
XX
DE   Stress tolerant plant-related transcription factor gene SeqID483.
XX
KW   transcription factor; transgenic plant; agriculture; drought resistance;
KW   stress tolerance; gene; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO2005047516-A2.
XX
PD   26-MAY-2005.
XX
PF   12-NOV-2004; 2004WO-US037584.
XX
PR   13-NOV-2003; 2003US-00714887.
PR   05-DEC-2003; 2003US-0527658P.
PR   05-FEB-2004; 2004US-0542928P.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
PI   Heard JE,  Riechmann JL,  Creelman RA,  Ratcliffe OJ,  Canales RD;
PI   Repetti P,  Kumimoto RW,  Gutterson NI,  Reuber TL,  Pineda O;
PI   Sherman BK,  Morrison TA,  Keddie JS,  Jiang C,  Century KS,  Adam L;
PI   Zhang JZ,  Hempel FD,  Libby JM;
XX
DR   WPI; 2005-372386/38.
DR   P-PSDB; AEA26643.
XX
PT   New transgenic plants for producing commercially or agriculturally useful
PT   plants having improved tolerance to drought, shade and low nitrogen
PT   conditions.
XX
PS   Disclosure; SEQ ID NO 483; 407pp; English.
XX
CC   This invention relates to a novel plant transcription factor
CC   polypeptides, the DNA sequences which encode them and their use in
CC   creating transgenic plants. The transgenic plant and methods are useful
CC   for producing commercially or agriculturally useful plants having
CC   improved tolerance to drought, shade and low nitrogen conditions when
CC   compared to wild-type reference plants. The present sequence is that of a
CC   plant transcription factor gene which was used during the development of
CC   the transgenic plants of the invention.
XX
SQ   Sequence 2930 BP; 906 A; 582 C; 692 G; 750 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              10.5           Length:        2930
Score:                  82.00          Matches:       33
Percent Similarity:     45.38%         Conservative:  21
Best Local Similarity:  27.73%         Mismatches:    39
Query Match:            14.80%         Indels:        26
DB:                     14             Gaps:          6

US-09-837-344-19 (1-107) x AEA26642 (1-2930)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGlu 21
              |||         ||||||   :::          |||   |||:::
Db        412 AAAGATGGGATAAAGAAACAGATGACCCCACCTGCAGAGGAGGGTAAT------------ 459

Qy         22 AspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGly 41
                        :::   |||         :::|||   |||:::|||:::::::::|||
Db        460 ------GGTCAGAAGTTTGAAGACGCACCTGTTGAAACACCGAACGAGGATAAAAAAGGA 513

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArg---Asp 60
                           :::|||||||||:::::::::     :::|||:::   |||
Db        514 ---------GATGGTGATGCAAGCTTGCCCAAAGAGGATGAGAGCTCAAGCAAACAAGAT 564

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGlu---------------------- 73
              ::::::|||           |||||||||      |||
Db        565 AACCAGGAAGAGAAGAAAGAAGAGAAAACTAAAGAAGAGTTTACTCCATCCAGTGAGACG 624

Qy         74 ---SerIleThrThrAsnValGluGlyArgArgAspIleHisLys--------------- 87
                 |||   |||         |||   ::::::|||      |||
Db        625 AAAAGTGAGACCGAAGGTGGGGAAGATCAAAAAGATGACTCCAAGTCAGAAAATGGTGGT 684

Qy         88 ---GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAsp 105
                 |||   |||:::|||||||||||||      :::|||           |||:::
Db        685 GGAGGTGATTTAGATGAGAAAAAAGAT------TTGAAAGACAACTCAGATGAAGAA 735

RESULT 10
AAH84231
```

ID AAH84231 standard; cDNA; 696 BP.
XX
AC AAH84231;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717 ORF5, SEQ:193.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
KW open reading frame; ORF; ss.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033547.
XX
PR 14-DEC-1999; 99US-00461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
DR WPI; 2001-390297/41.
DR P-PSDB; AAG98701.
XX
PT Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders.
XX
PS Claim 2; Fig 9E; 325pp; English.
XX
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC protective ORFs. The cell death protective cDNA clones are able to
CC prevent, delay or reverse progression through the apoptotic or necrotic
CC pathways when injected into a cell predisposed to or undergoing cell
CC death. The cell death protective nucleic acids and polypeptides can be
CC used in the diagnosis and treatment of disorders associated with cell
CC death, and to screen for compounds which modulate their activity or
CC expression. Such modulators, preferably a small organic molecule, an
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC cell death-related diseases. Such diseases include those associated with
CC the central nervous system including psychiatric or neurological
CC disorders, especially ischaemia-related conditions such as strokes, and
CC also includes neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, or Parkinson's disease. The modulators may also be
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC ; eye conditions such as diabetic retinopathy or macular degeneration;
CC hypertension; myocardial infarction; atherosclerosis; respiratory
CC conditions such as asthma or chronic obstructive pulmonary disease;
CC neoplastic conditions such as cancers or benign tumours; blood cell
CC conditions such as anaemia; gastrointestinal conditions such as gastritis
CC or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC conditions such as dermatitis or urticaria; or immune system disorders
CC such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC additionally be used to generate animal models of cell death-associated
CC disorders. The present sequence represents a cell death protective ORF
XX
SQ Sequence 696 BP; 368 A; 43 C; 211 G; 74 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.82	Length:	696
Score:	81.50	Matches:	31
Percent Similarity:	44.34%	Conservative:	16
Best Local Similarity:	29.25%	Mismatches:	54
Query Match:	14.71%	Indels:	5
DB:	4	Gaps:	3

US-09-837-344-19 (1-107) x AAH84231 (1-696)

```
Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        232 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 291

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||      :::   ||||||:::|||
Db        292 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 351

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                    :::                 :::|||   :::||||||      |||
Db        352 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 411

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||           ||||||       :::         |||   |||
Db        412 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 471

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||      :::
Db        472 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 531

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        532 AAAGAAGATGAAAAAGAG 549
```

RESULT 11
AAH84230
ID AAH84230 standard; cDNA; 699 BP.
XX
AC AAH84230;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717 ORF4, SEQ:191.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
KW open reading frame; ORF; ss.
XX

```
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   P-PSDB; AAG98700.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 2; Fig 9D; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective ORF
XX
SQ   Sequence 699 BP; 369 A; 43 C; 212 G; 75 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.83          Length:         699
Score:                  81.50         Matches:        31
Percent Similarity:     44.34%        Conservative:   16
Best Local Similarity:  29.25%        Mismatches:     54
Query Match:            14.71%        Indels:         5
DB:                     4             Gaps:           3

US-09-837-344-19 (1-107) x AAH84230 (1-699)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         |||||
Db        235 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 294

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||          |||       :::   ||||||:::|||
Db        295 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 354

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::              :::|||   :::||||||      |||
Db        355 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 414

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::         |||   |||
Db        415 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 474

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||          ||||||     :::|||   ||||||      :::
Db        475 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 534

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        535 AAAGAAGATGAAAAAGAG 552

RESULT 12
AAH84229
ID   AAH84229 standard; cDNA; 717 BP.
XX
AC   AAH84229;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human cell death protective cDNA clone CNI-00717 ORF3, SEQ:189.
XX
KW   Cell death protective; apoptosis; necrosis; human; drug screening;
KW   cell death-associated disorder; central nervous system disorder;
KW   psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW   stroke; cerebral infarction; ischaemic encephalopathy;
KW   neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW   Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW   vascular disease; opthalmological disorder; diabetic retinopathy;
KW   macular degeneration; hypertension; myocardial infarction;
KW   atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW   chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW   benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW   ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW   glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW   Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW   urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
KW   open reading frame; ORF; ss.
XX
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   P-PSDB; AAG98699.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 2; Fig 9C; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
```

```
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective ORF
XX
SQ   Sequence 717 BP; 379 A; 44 C; 216 G; 78 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.9          Length:         717
Score:                  81.50        Matches:        31
Percent Similarity:     44.34%       Conservative:   16
Best Local Similarity:  29.25%       Mismatches:     54
Query Match:            14.71%       Indels:         5
DB:                     4            Gaps:           3

US-09-837-344-19 (1-107) x AAH84229 (1-717)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        253 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 312

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                 :::   |||           |||       :::   ||||||:::|||
Db        313 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 372

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                      :::                 :::|||   :::||||||      |||
Db        373 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 432

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||             ||||||       :::          |||   |||
Db        433 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 492

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||             ||||||      :::|||   ||||||      :::
Db        493 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 552

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        553 AAAGAAGATGAAAAAGAG 570

RESULT 13
AAH84228
ID   AAH84228 standard; cDNA; 774 BP.
XX
AC   AAH84228;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human cell death protective cDNA clone CNI-00717 ORF2, SEQ:187.
XX
KW   Cell death protective; apoptosis; necrosis; human; drug screening;
KW   cell death-associated disorder; central nervous system disorder;
KW   psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW   stroke; cerebral infarction; ischaemic encephalopathy;
KW   neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW   Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW   vascular disease; opthalmological disorder; diabetic retinopathy;
KW   macular degeneration; hypertension; myocardial infarction;
KW   atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW   chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW   benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW   ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW   glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW   Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW   urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
KW   open reading frame; ORF; ss.
XX
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   P-PSDB; AAG98698.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 2; Fig 9B; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
```

```
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective ORF
XX
SQ   Sequence 774 BP; 401 A; 55 C; 229 G; 89 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.1          Length:        774
Score:                  81.50        Matches:       31
Percent Similarity:     44.34%       Conservative:  16
Best Local Similarity:  29.25%       Mismatches:    54
Query Match:            14.71%       Indels:        5
DB:                     4            Gaps:          3

US-09-837-344-19 (1-107) x AAH84228 (1-774)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        310 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 369

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||      :::   ||||||:::|||
Db        370 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 429

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                     :::                 :::|||   :::||||||      |||
Db        430 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 489

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||          ||||||        :::         |||   |||
Db        490 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 549

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||      :::
Db        550 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 609

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        610 AAAGAAGATGAAAAAGAG 627

RESULT 14
AAH84227
ID   AAH84227 standard; cDNA; 819 BP.
XX
AC   AAH84227;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Human cell death protective cDNA clone CNI-00717 ORF1, SEQ:185.
XX
KW   Cell death protective; apoptosis; necrosis; human; drug screening;
KW   cell death-associated disorder; central nervous system disorder;
KW   psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW   stroke; cerebral infarction; ischaemic encephalopathy;
KW   neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW   Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW   vascular disease; opthalmological disorder; diabetic retinopathy;
KW   macular degeneration; hypertension; myocardial infarction;
KW   atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW   chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW   benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW   ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW   glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW   Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW   urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
KW   open reading frame; ORF; ss.
XX
OS   Homo sapiens.
XX
PN   WO200145638-A2.
XX
PD   28-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033547.
XX
PR   14-DEC-1999;   99US-00461697.
XX
PA   (COGE-) COGENT NEUROSCIENCE INC.
XX
PI   Lo DC,  Barney S,  Thomas MB,  Portbury SD,  Puranam K,  Katz LC;
XX
DR   WPI; 2001-390297/41.
DR   P-PSDB; AAG98697.
XX
PT   Novel protective sequence polynucleotides and polypeptides, used to
PT   identify modulators of their expression and activity, which are used in
PT   to treat central nervous system conditions, diseases and disorders.
XX
PS   Claim 2; Fig 9A; 325pp; English.
XX
CC   Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC   protect against cell death (i.e., apoptosis or necrosis). Sequences
CC   AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC   AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC   while the remaining nucleic acid sequences within the range given above
CC   represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC   AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC   protective ORFs. The cell death protective cDNA clones are able to
CC   prevent, delay or reverse progression through the apoptotic or necrotic
CC   pathways when injected into a cell predisposed to or undergoing cell
CC   death. The cell death protective nucleic acids and polypeptides can be
CC   used in the diagnosis and treatment of disorders associated with cell
CC   death, and to screen for compounds which modulate their activity or
CC   expression. Such modulators, preferably a small organic molecule, an
CC   antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC   cell death-related diseases. Such diseases include those associated with
CC   the central nervous system including psychiatric or neurological
CC   disorders, especially ischaemia-related conditions such as strokes, and
CC   also includes neurodegenerative disorders such as Alzheimer's disease,
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a cell death protective ORF
XX
SQ   Sequence 819 BP; 417 A; 65 C; 240 G; 97 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.26         Length:        819
Score:                  81.50        Matches:       31
Percent Similarity:     44.34%       Conservative:  16
Best Local Similarity:  29.25%       Mismatches:    54
Query Match:            14.71%       Indels:        5
DB:                     4            Gaps:          3

US-09-837-344-19 (1-107) x AAH84227 (1-819)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        355 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 414

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
```

```
                    :::   |||             |||      :::   ||||||:::|||
Db        415 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 474

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                  :::|||   :::||||||     |||
Db        475 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 534

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||          ||||||      :::      |||   |||
Db        535 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 594

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||             ||||||      :::|||   ||||||      :::
Db        595 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 654

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        655 AAAGAAGATGAAAAAGAG 672
```

RESULT 15
ABN66837
ID ABN66837 standard; DNA; 1137 BP.
XX
AC ABN66837;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 1587.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
DR P-PSDB; ABP26206.
XX
PT New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
PS Claim 7; Page 3309; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 1137 BP; 468 A; 172 C; 208 G; 289 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.48	Length:	1137
Score:	81.50	Matches:	22
Percent Similarity:	51.95%	Conservative:	18
Best Local Similarity:	28.57%	Mismatches:	20
Query Match:	14.71%	Indels:	17
DB:	6	Gaps:	3

US-09-837-344-19 (1-107) x ABN66837 (1-1137)

```
Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db        715 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 774

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||::::::::::|||:::|||      :::           |||:::
Db        775 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 834

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                                  |||:::      |||       |||::::::||||||
Db        835 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGGTCAGAC 894

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db        895 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 927
```

RESULT 16
ADU69448
ID ADU69448 standard; DNA; 1239 BP.
XX
AC ADU69448;
XX
DT 10-FEB-2005 (first entry)
XX
DE S agalactiae hyperimmune serum reactive antigen DNA seqid 143.
XX
KW immune stimulation; antigen; bacterial surface display;
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;
KW antibacterial; infection; gene; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO2004099242-A2.
XX
PD 18-NOV-2004.
XX
PF 06-MAY-2004; 2004WO-EP004856.
XX
PR 07-MAY-2003; 2003EP-00450112.
PR 28-NOV-2003; 2003EP-00450266.
XX
PA (INTE-) INTERCELL AG.
XX
PI Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;
XX
DR WPI; 2004-821662/81.
DR P-PSDB; ADU69665.
XX
PT New nucleic acid molecule encoding a hyperimmune serum reactive antigen,
PT useful for the manufacture of a vaccine against Streptococcus agalactiae
PT infection.
XX
PS Claim 3; SEQ ID NO 143; 221pp; English.
XX
CC The invention describes an isolated nucleic acid molecule encoding a
CC hyperimmune serum reactive antigen or its fragment. Also described are: a

```
CC   vector comprising the nucleic acid molecule; a host cell comprising the
CC   vector; a hyperimmune serum-reactive antigen comprising a sequence
CC   encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467
CC   or 812 amino acids; fragments of hyperimmune serum-reactive antigens
CC   consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;
CC   a process for producing a Streptococcus agalactiae hyperimmune serum
CC   reactive antigen; a process for producing a cell that expresses a S.
CC   agalactiae hyperimmune serum reactive antigen; a pharmaceutical
CC   composition, especially a vaccine, comprising the hyperimmune serum-
CC   reactive antigen or nucleic acid molecule; an antibody that binds at
CC   least to a selective part of the hyperimmune serum-reactive antigen; a
CC   hybridoma cell line, which produces the antibody; a method for producing
CC   the antibody; an antagonist that binds to the hyperimmune serum-reactive
CC   antigen; a method for identifying an antagonist capable of binding to the
CC   hyperimmune serum-reactive antigen; a method for identifying an
CC   antagonist capable of reducing or inhibiting the interaction activity of
CC   a hyperimmune serum-reactive antigen to its interaction partner; a
CC   process for in vitro diagnosing a disease related to expression of the
CC   hyperimmune serum-reactive antigen; and a process for in vitro diagnosis
CC   of a bacterial infection, especially a S. agalactiae infection. The
CC   hyperimmune serum reactive antigen is useful for isolating, purifying
CC   and/or identifying an interaction partner of the hyperimmune serum
CC   reactive antigen. The hyperimmune serum reactive antigen is useful for
CC   generating a peptide binding to the hyperimmune serum reactive antigen,
CC   where the peptide comprises anticalines, or for the manufacture of a
CC   functional nucleic acid comprising aptamers or spiegelmers. The nucleic
CC   acid molecule is useful for the manufacture of a functional ribonucleic
CC   acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic
CC   acid molecule, hyperimmune serum-reactive antigen or antibody is useful
CC   for the manufacture of a vaccine against S. agalactiae infection. This
CC   sequence represents a Streptococcus agalactiae hyperimmune serum reactive
CC   antigen.
XX
SQ   Sequence 1239 BP; 510 A; 186 C; 225 G; 318 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.89          Length:        1239
Score:                  81.50         Matches:       22
Percent Similarity:     51.95%        Conservative:  18
Best Local Similarity:  28.57%        Mismatches:    20
Query Match:            14.71%        Indels:        17
DB:                     13            Gaps:          3

US-09-837-344-19 (1-107) x ADU69448 (1-1239)

Qy          42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
               ||||||:::   |||   :::   ||||||::::::
Db         814 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 873

Qy          54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                  |||:::|||::::::::::|||:::|||      :::            |||:::
Db         874 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 933

Qy          71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                           |||:::     |||     |||::::::|||||
Db         934 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 993

Qy          91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
               ||||||                 :::||||||::::::||||||:::|||
Db         994 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 1026

RESULT 17
ADV85196
ID   ADV85196 standard; DNA; 1239 BP.
XX
AC   ADV85196;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID 6337.
XX
KW   Antibacterial; vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200292818-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P,  Rusniok C,  Chevalier F,  Frangeul L,  Lalioui L;
PI   Zouine M,  Couve E,  Buchrieser C,  Poyart C,  Trieu-Cuot P,  Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
PS   Claim 4; SEQ ID NO 6337; 439pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC   novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC   nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC   synthesis of amino acids, cell membranes, intermediate (central)
CC   metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC   nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC   regulatory functions, replication, transcription, translation, protein
CC   transport, adaptation to atypical conditions, sensitivity to medicines
CC   and/or analogues, functions related to transposons, biosynthesis of
CC   cofactors, prosthetic groups and transporters, cell membrane proteins and
CC   cellular machinery. (I) are useful for the detection and/or amplification
CC   of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC   useful for treatment of a bacterial S. agalactiae infection. The complete
CC   genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC   present patent is an equivalent for the basic patent FR2824074A1, which
CC   contains only 2344 sequences.
XX
SQ   Sequence 1239 BP; 510 A; 186 C; 225 G; 318 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.89          Length:        1239
Score:                  81.50         Matches:       22
Percent Similarity:     51.95%        Conservative:  18
Best Local Similarity:  28.57%        Mismatches:    20
Query Match:            14.71%        Indels:        17
DB:                     13            Gaps:          3

US-09-837-344-19 (1-107) x ADV85196 (1-1239)

Qy          42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
               ||||||:::   |||   :::   ||||||::::::
Db         814 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 873

Qy          54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                  |||:::|||::::::::::|||:::|||      :::            |||:::
Db         874 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 933

Qy          71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                           |||:::     |||     |||::::::|||||
Db         934 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 993

Qy          91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
               ||||||                 :::||||||::::::||||||:::|||
Db         994 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 1026

RESULT 18
AAA05838
ID   AAA05838 standard; DNA; 1455 BP.
```

XX
AC AAA05838;
XX
DT 30-MAY-2000 (first entry)
XX
DE Group B Streptococcus protein encoding nucleotide sequence SEQ ID NO:38.
XX
KW Group B Streptococcus; Streptococcus agalactiae; protein antigen;
KW vaccine; screening; immunogen; detection; diagnosis; infection; antibody;
KW affibody; antibacterial; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200006736-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB002444.
XX
PR 27-JUL-1998; 98GB-00016335.
PR 19-MAR-1999; 99US-0125163P.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB;
XX
DR WPI; 2000-195299/17.
DR P-PSDB; AAY91310.
XX
PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of
PT Streptococcal infections and for screening of antibodies or affibodies.
XX
PS Claim 4; Fig 1; 123pp; English.
XX
CC AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given in
CC AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also
CC known as Streptococcus agalactiae. The GBS polynucleotides and
CC polypeptides have antibacterial activity. Immunogenic compositions
CC comprising GBS polynucleotides or polypeptides can be used as vaccines
CC and for the treatment or prophylaxis of GBS infection. The
CC polynucleotides and polypeptides can also be used in the detection of GBS
CC and for screening DNA encoding bacterial cell envelope associated or
CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
CC represent primers used in the exemplification of the present invention
XX
SQ Sequence 1455 BP; 584 A; 237 C; 271 G; 363 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.81	Length:	1455
Score:	81.50	Matches:	22
Percent Similarity:	51.95%	Conservative:	18
Best Local Similarity:	28.57%	Mismatches:	20
Query Match:	14.71%	Indels:	17
DB:	3	Gaps:	3

US-09-837-344-19 (1-107) x AAA05838 (1-1455)

```
Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db       1123 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAGGGTTGGGGACACGACAGTGAT 1182

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||:::::::::|||:::|||      :::           |||:::
Db       1183 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 1242

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                           |||:::      |||      |||::::::||||||
Db       1243 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 1302

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                 :::||||||::::::||||||:::|||
Db       1303 GAAGAA----------------GAACCAGAAAAACAAGAAGATGAATCA 1335
```

RESULT 19
AAH84226
ID AAH84226 standard; cDNA; 1669 BP.
XX
AC AAH84226;
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00717, SEQ:184.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; opthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS; ss.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033547.
XX
PR 14-DEC-1999; 99US-00461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
DR WPI; 2001-390297/41.
DR P-PSDB; AAG98697, AAG98698, AAG98699, AAG98700, AAG98701, AAG98702,
DR AAG98703, AAG98704, AAG98705, AAG98706, AAG98707, AAG98708, AAG98709,
DR AAG98710, AAG98711, AAG98712, AAG98713, AAG98714, AAG98715, AAG98716,
DR AAG98717, AAG98718, AAG98719, AAG98720, AAG98721, AAG98722, AAG98723,
DR AAG98724, AAG98725, AAG98726, AAG98727, AAG98728, AAG98729, AAG98730,
DR AAG98731, AAG98732, AAG98733, AAG98734.
XX
PT Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders.
XX
PS Example; Fig 1F; 325pp; English.
XX
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC protective ORFs. The cell death protective cDNA clones are able to
CC prevent, delay or reverse progression through the apoptotic or necrotic
CC pathways when injected into a cell predisposed to or undergoing cell
CC death. The cell death protective nucleic acids and polypeptides can be
CC used in the diagnosis and treatment of disorders associated with cell
CC death, and to screen for compounds which modulate their activity or
CC expression. Such modulators, preferably a small organic molecule, an
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC cell death-related diseases. Such diseases include those associated with
CC the central nervous system including psychiatric or neurological
CC disorders, especially ischaemia-related conditions such as strokes, and
CC also includes neurodegenerative disorders such as Alzheimer's disease,

```
CC   Huntington's disease, or Parkinson's disease. The modulators may also be
CC   used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC   vascular diseases such as ischaemic encephalopathy or cerebral infarction
CC   ; eye conditions such as diabetic retinopathy or macular degeneration;
CC   hypertension; myocardial infarction; atherosclerosis; respiratory
CC   conditions such as asthma or chronic obstructive pulmonary disease;
CC   neoplastic conditions such as cancers or benign tumours; blood cell
CC   conditions such as anaemia; gastrointestinal conditions such as gastritis
CC   or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney
CC   disorders such as glomerulonephritis; cystitis; endometriosis; endocrine
CC   disorders such as Grave's disease or Hashimoto's thyroiditis; skin
CC   conditions such as dermatitis or urticaria; or immune system disorders
CC   such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   additionally be used to generate animal models of cell death-associated
CC   disorders. The present sequence represents a full-length cell death
CC   protective cDNA
XX
SQ   Sequence 1669 BP; 721 A; 177 C; 357 G; 414 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.76          Length:        1669
Score:                  81.50         Matches:       31
Percent Similarity:     44.34%        Conservative:  16
Best Local Similarity:  29.25%        Mismatches:    54
Query Match:            14.71%        Indels:        5
DB:                     4             Gaps:          3

US-09-837-344-19 (1-107) x AAH84226 (1-1669)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::       ||||||
Db        434 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 493

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||           |||      :::   ||||||:::|||
Db        494 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 553

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                    :::                   :::|||   :::||||||      |||
Db        554 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 613

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||           ||||||      :::         |||   |||
Db        614 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 673

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||           ||||||      :::|||   ||||||      :::
Db        674 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 733

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        734 AAAGAAGATGAAAAAGAG 751

RESULT 20
AAS74470
ID   AAS74470 standard; cDNA; 2460 BP.
XX
AC   AAS74470;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10274.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10283.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 10274; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2460 BP; 991 A; 381 C; 573 G; 515 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              9.6           Length:        2460
Score:                  81.50         Matches:       36
Percent Similarity:     43.08%        Conservative:  20
Best Local Similarity:  27.69%        Mismatches:    37
Query Match:            14.71%        Indels:        38
DB:                     5             Gaps:          6

US-09-837-344-19 (1-107) x AAS74470 (1-2460)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLys------LysGluHisGlyAsp------ 17
              ||||||:::||||||   |||   |||   |||      :::   |||   :::
Db       1877 AAAGCAGAAACAAAGACAAATACAGAAACCAAGAGAGTAAGAGCTCACACAGAAAAGAGA 1936

Qy         18 ----------------------------------------IleLeuAlaGluAspLeu 23
                                                      |||:::|||:::
Db       1937 ATTCTGAGAGTGAGAAAAGAATGTACTCTAAAAGTCGTGATCATAATAGCTCAAATAACA 1996

Qy         24 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 43
                ||:::      :::      |||::::::||||||:::      :::
Db       1997 GCAGGGAAAAAAAGGCTGATAGAGATCAAAGTCCCTTCTCAAAAATA-AAA--------- 2046

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg-----------------GlyAsn 57
                  |||||||||      ||||||:::                         |||
Db       2047 ---------CAAAGCAGT------CAGGACGATGAATTAAAGTCCTCCATGTTGAAAAAT 2091

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
                ||||   |||      |||   :::||||||   |||::::::|||
Db       2092 AAGGAGGATGAGAAGATCAGATCCTCAGTGGAAAAAGAAAACCAAAAATCA--------- 2142

Qy         78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
```

```
                  :::|||:::   ::::::|||              :::|||   |||
Db       2143 ------AAAGGTCAAGAAAATGTCCATGTACATGAAAAAAATAAAAAATTTGATCATGAA 2196

Qy         98 IleLysProGluGlnLysGluAspLysSer 107
                 |||         ||||||||||||
Db       2197 TCAAGCCCTGGAACAGATGAAGACAAAAGC 2226
```

RESULT 21
ADV87739/c
ID ADV87739 standard; DNA; 75248 BP.
XX
AC ADV87739;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae DNA sequence, SEQ ID 133.
XX
KW Antibacterial; Vaccine; bacterial infection; ds.
XX
OS Streptococcus agalactiae.
XX
PN FR2824074-A1.
XX
PD 31-OCT-2002.
XX
PF 26-APR-2001; 2001FR-00005642.
XX
PR 26-APR-2001; 2001FR-00005642.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
XX
DR WPI; 2004-101891/11.
XX
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 1; SEQ ID NO 133; 2687pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
CC agalactiae involved in the synthesis of amino acids, cell membranes,
CC intermediate (central) metabolism, energetic metabolism, fatty acid and
CC phospholipid metabolism, nucleotide metabolism including purines,
CC pyrimidines and/or nucleosides, regulatory functions, replication,
CC transcription, translation, protein transport, adaptation to atypical
CC conditions, sensitivity to medicines and/or analogues, functions related
CC to transposons, biosynthesis of cofactors, prosthetic groups and
CC transporters, cell membrane proteins and cellular machinery. (I) are
CC useful for the detection and/or amplification of nucleic acids.
CC Pharmaceutical composition comprising (I) or (II) are useful for
CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
CC equivalent for the present basic patent FR2824074A1. WO200292818A2
CC contains 6617 sequence whereas the present patent only contains 2344
CC sequences.
XX
SQ Sequence 75248 BP; 24346 A; 14447 C; 11608 G; 24847 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	866	Length:	75248
Score:	81.50	Matches:	22
Percent Similarity:	51.95%	Conservative:	18
Best Local Similarity:	28.57%	Mismatches:	20
Query Match:	14.71%	Indels:	17
DB:	13	Gaps:	3

US-09-837-344-19 (1-107) x ADV87739 (1-75248)

```
Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db      20795 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 20736

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||:::::::::|||:::|||      :::            |||:::
Db      20735 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 20676

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                           |||:::      |||      |||::::::||||||
Db      20675 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 20616

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                 :::||||||::::::||||||:::|||
Db      20615 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 20583
```

RESULT 22
ADV78992/c
ID ADV78992 standard; DNA; 75248 BP.
XX
AC ADV78992;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae DNA sequence, SEQ ID 133.
XX
KW Antibacterial; vaccine; bacterial infection; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200292818-A2.
XX
PD 21-NOV-2002.
XX
PF 26-APR-2002; 2002WO-IB003059.
XX
PR 26-APR-2001; 2001FR-00005642.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR WPI; 2004-101891/11.
XX
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 1; SEQ ID NO 133; 439pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.
XX
SQ Sequence 75248 BP; 24346 A; 14447 C; 11608 G; 24847 T; 0 U; 0 Other;

Alignment Scores:

```
Pred. No.:              866           Length:        75248
Score:                  81.50         Matches:       22
Percent Similarity:     51.95%        Conservative:  18
Best Local Similarity:  28.57%        Mismatches:    20
Query Match:            14.71%        Indels:        17
DB:                     13            Gaps:          3

US-09-837-344-19 (1-107) x ADV78992 (1-75248)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp----------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db      20795 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 20736

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||::::::::::|||:::|||      :::            |||:::
Db      20735 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 20676

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                          |||:::      |||     |||::::::||||||
Db      20675 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 20616

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db      20615 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 20583

RESULT 23
ABN71527_18/c
Continuation (19 of 22) of ABN71527 from base 1800001 (Streptococcus polynucleotide SEQ
WP Sequence split into 22 fragments  LOCUS ABN71527 Accession Abn71527
WP     Fragment Name      Begin     End
WP     ABN71527_00            1    110000
WP     ABN71527_01       100001    210000
WP     ABN71527_02       200001    310000
WP     ABN71527_03       300001    410000
WP     ABN71527_04       400001    510000
WP     ABN71527_05       500001    610000
WP     ABN71527_06       600001    710000
WP     ABN71527_07       700001    810000
WP     ABN71527_08       800001    910000
WP     ABN71527_09       900001   1010000
WP     ABN71527_10      1000001   1110000
WP     ABN71527_11      1100001   1210000
WP     ABN71527_12      1200001   1310000
WP     ABN71527_13      1300001   1410000
WP     ABN71527_14      1400001   1510000
WP     ABN71527_15      1500001   1610000
WP     ABN71527_16      1600001   1710000
WP     ABN71527_17      1700001   1810000
WP     ABN71527_18      1800001   1910000
WP     ABN71527_19      1900001   2010000
WP     ABN71527_20      2000001   2110000
WP     ABN71527_21      2100001   2155561

Alignment Scores:
Pred. No.:              1.43e+03      Length:        110000
Score:                  81.50         Matches:       22
Percent Similarity:     51.95%        Conservative:  18
Best Local Similarity:  28.57%        Mismatches:    20
Query Match:            14.71%        Indels:        17
DB:                     6             Gaps:          3

US-09-837-344-19 (1-107) x ABN71527_18 (1-110000)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp----------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db      89187 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 89128

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||::::::::::|||:::|||      :::            |||:::
Db      89127 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 89068

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                          |||:::      |||     |||::::::||||||
Db      89067 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGGTCAGAC 89008

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db      89007 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 88975

RESULT 24
ADV81204_19/c
Continuation (20 of 23) of ADV81204 from base 1900001 (Streptococcus agalactiae complete
WP Sequence split into 23 fragments  LOCUS ADV81204 Accession Adv81204
WP     Fragment Name      Begin     End
WP     ADV81204_00            1    110000
WP     ADV81204_01       100001    210000
WP     ADV81204_02       200001    310000
WP     ADV81204_03       300001    410000
WP     ADV81204_04       400001    510000
WP     ADV81204_05       500001    610000
WP     ADV81204_06       600001    710000
WP     ADV81204_07       700001    810000
WP     ADV81204_08       800001    910000
WP     ADV81204_09       900001   1010000
WP     ADV81204_10      1000001   1110000
WP     ADV81204_11      1100001   1210000
WP     ADV81204_12      1200001   1310000
WP     ADV81204_13      1300001   1410000
WP     ADV81204_14      1400001   1510000
WP     ADV81204_15      1500001   1610000
WP     ADV81204_16      1600001   1710000
WP     ADV81204_17      1700001   1810000
WP     ADV81204_18      1800001   1910000
WP     ADV81204_19      1900001   2010000
WP     ADV81204_20      2000001   2110000
WP     ADV81204_21      2100001   2210000
WP     ADV81204_22      2200001   2217924

Alignment Scores:
Pred. No.:              1.43e+03      Length:        110000
Score:                  81.50         Matches:       22
Percent Similarity:     51.95%        Conservative:  18
Best Local Similarity:  28.57%        Mismatches:    20
Query Match:            14.71%        Indels:        17
DB:                     13            Gaps:          3

US-09-837-344-19 (1-107) x ADV81204_19 (1-110000)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp----------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db     103361 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 103302

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||::::::::::|||:::|||      :::            |||:::
Db     103301 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 103242

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                          |||:::      |||     |||::::::||||||
Db     103241 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 103182

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db     103181 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 103149

RESULT 25
ADV81204_20/c
Continuation (21 of 23) of ADV81204 from base 2000001 (Streptococcus agalactiae complete
WP Sequence split into 23 fragments  LOCUS ADV81204 Accession Adv81204
WP     Fragment Name      Begin     End
WP     ADV81204_00            1    110000
WP     ADV81204_01       100001    210000
WP     ADV81204_02       200001    310000
WP     ADV81204_03       300001    410000
WP     ADV81204_04       400001    510000
WP     ADV81204_05       500001    610000
WP     ADV81204_06       600001    710000
```

```
WP      ADV81204_07        700001    810000
WP      ADV81204_08        800001    910000
WP      ADV81204_09        900001   1010000
WP      ADV81204_10       1000001   1110000
WP      ADV81204_11       1100001   1210000
WP      ADV81204_12       1200001   1310000
WP      ADV81204_13       1300001   1410000
WP      ADV81204_14       1400001   1510000
WP      ADV81204_15       1500001   1610000
WP      ADV81204_16       1600001   1710000
WP      ADV81204_17       1700001   1810000
WP      ADV81204_18       1800001   1910000
WP      ADV81204_19       1900001   2010000
WP      ADV81204_20       2000001   2110000
WP      ADV81204_21       2100001   2210000
WP      ADV81204_22       2200001   2217924

Alignment Scores:
Pred. No.:              1.43e+03        Length:         110000
Score:                  81.50           Matches:        22
Percent Similarity:     51.95%          Conservative:   18
Best Local Similarity:  28.57%          Mismatches:     20
Query Match:            14.71%          Indels:         17
DB:                     13              Gaps:           3

US-09-837-344-19 (1-107) x ADV81204_20 (1-110000)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
              ||||||:::   |||   :::   ||||||::::::
Db       3361 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAAGGTTGGGGACACGACAGTGAT 3302

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                 |||:::|||::::::::::|||:::|||      :::            |||:::
Db       3301 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 3242

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                          |||:::      |||      |||::::::||||||
Db       3241 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 3182

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db       3181 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 3149

RESULT 26
ADX11745
ID   ADX11745 standard; cDNA; 1508 BP.
XX
AC   ADX11745;
XX
DT   21-APR-2005  (first entry)
XX
DE   Plant full length insert polynucleotide seqid 6320.
XX
KW   plant protectant; plant growth regulant; gene therapy; plant;
KW   recombinant DNA construct; physical array; plant breeding marker;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW   extreme osmotic condition; pathogen tolerance; pest tolerance;
KW   growth rate; cell cycle pathway; disease resistance;
KW   galactomannan production; lignin production; plant growth regulator;
KW   yield; plant growth; plant development; seed oil; protein yield;
KW   protein content; gene; ss.
XX
OS   Unidentified.
XX
PN   US2004034888-A1.
XX
PD   19-FEB-2004.
XX
PF   28-APR-2003; 2003US-00425114.
XX
PR   06-MAY-1999;   99US-00304517.
PR   05-NOV-2001; 2001US-00985678.
XX
PA   (LIUJ/) LIU J.
PA   (ZHOU/) ZHOU Y.
PA   (KOVA/) KOVALIC D K.
PA   (SCRE/) SCREEN S E.
PA   (TABA/) TABASKA J E.
PA   (CAOY/) CAO Y.
XX
PI   Liu J,  Zhou Y,  Kovalic DK,  Screen SE,  Tabaska JE,  Cao Y;
XX
DR   WPI; 2004-180133/17.
XX
PT   New recombinant DNA construct, useful for improving plant tolerance to
PT   cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT   pests, for conferring increased resistance to plant disease, or for
PT   improving yield.
XX
PS   Claim 1; SEQ ID NO 6320; 15pp; English.
XX
CC   The invention describes a recombinant DNA construct comprising a
CC   polynucleotide consisting of a sequence encoding an amino acid sequence
CC   available in electronic form from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC   of the invention are also useful in physical arrays of molecules and as
CC   plant breeding markers. The recombinant DNA construct is useful for
CC   improving plant tolerance to cold, heat, drought, herbicides, extreme
CC   osmotic conditions, pathogens or pests, for manipulating growth rate in
CC   plant cells by modification of the cell cycle pathway, for conferring
CC   increased resistance to plant disease, for producing galactomannan,
CC   lignin or plant growth regulators, for increasing the rate of homologous
CC   recombination in plants, for improving yield by modification of
CC   photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC   or by providing improved plant growth and development under at least one
CC   stress condition or for modifying seed oil or protein yield and/or
CC   content. This sequence represents a plant full length insert
CC   polynucleotide that can be used in the recombinant DNA construct of the
CC   invention.
XX
SQ   Sequence 1508 BP; 497 A; 236 C; 346 G; 429 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.82            Length:         1508
Score:                  81.00           Matches:        26
Percent Similarity:     44.64%          Conservative:   24
Best Local Similarity:  23.21%          Mismatches:     32
Query Match:            14.62%          Indels:         30
DB:                     13              Gaps:           5

US-09-837-344-19 (1-107) x ADX11745 (1-1508)

Qy          4 AspThrLysLysAsnLeuGlu--------------------------ArgLysLysGlu 14
              ||||||          :::|||                          ||||||:::
Db       1056 GATACAGCGATGGAAATAGAAGCACCAGTTGAAGTTTCAGGCAAGAAAGCCAAGAAGAAG 1115

Qy         15 HisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeu 34
              |||      ::::::   :::|||:::                        |||::::::
Db       1116 CATACAGCTGTCGTTACAGATGACATG--------------------GCTGTAGATAAG 1154

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn 54
              |||:::|||          |||      :::                  :::|||:::
Db       1155 CCCACAGAGACTACAAATGGTGATGCATTG--------------------GAGGATCAT 1193

Qy         55 ArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSer 74
              :::          :::      |||                  |||      |||:::
Db       1194 AAGTCAGAAAAGAAGAAGAGAAAG-----------------AAAGAAAAGAGGAAATTG 1235

Qy         75 IleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLys 94
                 |||   |||   :::         |||         |||   :::|||   ::::::
Db       1236 GACACGGAGAACGATCAAGCCATGGATGATGGTGCTAATGGG---GTTGAAAGCGAACAA 1292

Qy         95 AspGlySerIleLysProGluGlnLysGluAspLys 106
              ||||||::::::|||   ::::::|||:::||||||
Db       1293 GATGGAGCAGTTAAAAAGAAGAAAAAAAAGGATAAA 1328
```

```
RESULT 27
ABV49476
ID   ABV49476 standard; cDNA; 540 BP.
XX
AC   ABV49476;
XX
DT   17-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 49467.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 9661; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 540 BP; 248 A; 85 C; 120 G; 87 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.74          Length:         540
Score:                  80.50         Matches:        37
Percent Similarity:     42.31%        Conservative:   18
Best Local Similarity:  28.46%        Mismatches:     38
Query Match:            14.53%        Indels:         38
DB:                     5             Gaps:           6

US-09-837-344-19 (1-107) x ABV49476 (1-540)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLys------LysGluHisGlyAsp------ 17
              ||||||:::||||||   |||   |||   |||      :::   |||   :::
Db        167 AAAGCAGAAACAAAGACAAATACAGAAACCAAGAGAGTAAGAGCTCACACAGAAAAGAAA 226

Qy         18 ----------------------------------------IleLeuAlaGluAspLeu 23
                                                      |||:::|||:::
Db        227 ATTCTGAGAGTGAGAAAAGAATGTACTCTAAAAGTCGTGATCATAATAGCTCAAATAACA 286

Qy         24 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 43
               |||:::      :::      |||::::::||||||:::      :::
Db        287 GCAGGGAAAAAAAGGCTGATAGAGATCAAAGTCCCTTCTCAAAAATA-AAA--------- 336

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg-----------------GlyAsn 57
                 |||||||||      |||||||||                          |||
Db        337 ---------CAAAGCAGT------CAGGACAATGAATTAAAGTCCTCCATGTTGAAAAAT 381

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
               |||   |||       |||   :::||||||   |||:::::::|||
Db        382 AAGGAGGATGAGAAGATCAGATCCTCAGTGGAAAAAGAAAACCAAAAATCA--------- 432

Qy         78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
                 :::|||:::   :::   |||                :::|||   |||
Db        433 ------AAAGGTCAAGAAAATGACCATGTACATGAAAAAAATAAAAAATTTGATCATGAA 486

Qy         98 IleLysProGluGlnLysGluAspLysSer 107
               |||         ||||||||||||
Db        487 TCAAGCCCTGGAACAGATGAAGACAAAAGC 516

RESULT 28
ADZ72254
ID   ADZ72254 standard; DNA; 1116 BP.
XX
AC   ADZ72254;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Plasmodium falciparum MSP-3-like gene, MSP-3-2 SEQ ID NO: 3.
XX
KW   Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;
KW   infection; merozoite surface protein 3-like gene; MSP-3-2; chromosome 10;
KW   gene; ds.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1116
FT                   /*tag=  a
FT                   /product= "Plasmodium falciparum MSP-3-2 protein"
XX
PN   EP1526178-A1.
XX
PD   27-APR-2005.
XX
PF   24-OCT-2003; 2003EP-00292673.
XX
PR   24-OCT-2003; 2003EP-00292673.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Druilhe P;
XX
DR   WPI; 2005-323987/34.
DR   P-PSDB; ADZ72255.
XX
PT   Novel MSP-3-like family genes located on chromosome 10 of Plasmodium
PT   falciparum, which encode proteins useful for preparing vaccine
PT   compositions against malaria.
XX
PS   Claim 6; SEQ ID NO 3; 137pp; English.
XX
CC   The present invention relates to the protection against malaria. More
CC   particularly, the invention pertains to a family of MSP-3 (merozoite
CC   surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-
CC   5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium
CC   falciparum, highly conserved in P. falciparum strains, simultaneously
CC   expressed in P. falciparum at the erythrocytic stages and encoding
CC   proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at
CC   their N-terminal extremity and which are located at the merozoite
CC   surface. The characterization of this gene family enables the definition
CC   of immunogenic and vaccine compositions against P. falciparum. The
CC   present sequence is the P. falciparum MSP-3-2 gene.
```

```
XX
SQ   Sequence 1116 BP; 536 A; 117 C; 184 G; 279 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.53          Length:        1116
Score:                  80.50         Matches:       29
Percent Similarity:     46.88%        Conservative:  16
Best Local Similarity:  30.21%        Mismatches:    38
Query Match:            14.53%        Indels:        13
DB:                     14            Gaps:          4

US-09-837-344-19 (1-107) x ADZ72254 (1-1116)

Qy         17 AspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGlu--------- 33
              |||         |||   |||      :::::::::||||||:::   :::
Db        595 GATAAAAAGACAGAATATTTACTAGAACAAATAAAAATTCCATCATGGGATAGAAATAAC 654

Qy         34 LeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAsp 53
              :::|||   |||||||||:::          |||                  :::|||
Db        655 ATCCCCGATGAGAATGAACAAGTAATAGAGGACCCTCAA--------------GAAGAT 699

Qy         54 AsnArgGlyAsnSerArgAspSerLys---------GluIleSerIleIleGluLysThr 70
              |||:::             |||   :::         |||             |||
Db        700 AATAAAGATGAAGATGAAGATGAAGAAACAGAAACAGAAAATTTGGAAACAGAAGATGAT 759

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
              |||   |||   |||      |||   |||        ||||||   :::     :::
Db        760 AATAATGAAGAGATAGAAGAAAATGAAGAAGAT------GACATAGATGAAGAAAGTGTA 813

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
              |||||||||:::::         |||   |||:::::|||   |||
Db        814 GAAGAAAAGGAAGAAGAGGAAGAAAAAAAGGAAGAAGAAGAAAAAAAG 861

RESULT 29
AAV74832
ID   AAV74832 standard; DNA; 1353 BP.
XX
AC   AAV74832;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #521.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    301. .360
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They are
FT                   included to maintain the nucleotide numbering given in
FT                   the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-00100117.
XX
PR   05-JAN-1996;   96US-0009861P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA,  Choi GH,  Barash SC,  Dillon PJ,  Fannon MR,  Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of anti-
PT   S.aureus vaccines.
XX
PS   Claim 1; Page 1450-1451; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the computer
CC   readable medium
XX
SQ   Sequence 1353 BP; 578 A; 165 C; 219 G; 330 T; 0 U; 61 Other;

Alignment Scores:
Pred. No.:              5.83          Length:        1353
Score:                  80.50         Matches:       30
Percent Similarity:     45.38%        Conservative:  24
Best Local Similarity:  25.21%        Mismatches:    39
Query Match:            14.53%        Indels:        26
DB:                     2             Gaps:          4

US-09-837-344-19 (1-107) x AAV74832 (1-1353)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        461 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 520

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::      ::::::                ||||||   :::|||
Db        521 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 580

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::             |||                  :::
Db        581 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 640

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                  ||| |||:::     :::|||:::   ||||||:::   |||::::::
Db        641 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 700

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                :                  :::|||   |||
Db        701 CACAAAC-----------------------------------CTGCATCTCAAGAA 721

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 |||      ||||||          ||||||   ::::::||||||   :::
Db        722 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 776

RESULT 30
ACF73779
ID   ACF73779 standard; DNA; 1407 BP.
XX
AC   ACF73779;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Staphylococcus aureus DNA #1459.
XX
KW   Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW   enzymatic assay; antibiotic target; gene; ds.
XX
```

OS Staphylococcus aureus.
XX
PN WO200294868-A2.
XX
PD 28-NOV-2002.
XX
PF 27-MAR-2002; 2002WO-IB002637.
XX
PR 27-MAR-2001; 2001GB-00007661.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Masignani V, Mora M, Scarselli M;
XX
DR WPI; 2003-120786/11.
DR P-PSDB; ABM72219.
XX
PT New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
PS Claim 6; SEQ ID NO 2917; 49pp; English.
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus genes of the invention
XX
SQ Sequence 1407 BP; 641 A; 172 C; 237 G; 357 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.14	Length:	1407
Score:	80.50	Matches:	30
Percent Similarity:	45.38%	Conservative:	24
Best Local Similarity:	25.21%	Mismatches:	39
Query Match:	14.53%	Indels:	26
DB:	8	Gaps:	4

US-09-837-344-19 (1-107) x ACF73779 (1-1407)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        514 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 573

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::      ::::::                 ||||||   :::|||
Db        574 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 633

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::             |||                  :::
Db        634 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 693

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                  ||| |||:::      :::|||:::   ||||||:::   |||::::::
Db        694 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 753

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                                      :::|||   |||
Db        754 CACAAAC-----------------------------------CTGCATCTCAAGAA 774

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 |||      ||||||         ||||||   ::::::||||||   :::
Db        775 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 829
```

Search completed: December 20, 2005, 01:36:05
Job time : 596 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 02:40:33 ; Search time 166 Seconds
(without alignments)
1145.778 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103707_224/app_query.fasta_1.263
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_193_@runat_15122005_103707_224 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	950	3	US-08-098-327E-32	Sequence 32, Appl
2	554	100.0	950	3	US-08-462-625-32	Sequence 32, Appl
3	554	100.0	1482	3	US-08-098-327E-41	Sequence 41, Appl
4	554	100.0	1482	3	US-08-462-625-41	Sequence 41, Appl
5	554	100.0	1493	3	US-08-098-327E-38	Sequence 38, Appl
6	554	100.0	1493	3	US-08-462-625-38	Sequence 38, Appl
7	553	99.8	1320	2	US-08-257-073-15	Sequence 15, Appl
8	551	99.5	1482	3	US-08-098-327E-45	Sequence 45, Appl
9	551	99.5	1482	3	US-08-462-625-45	Sequence 45, Appl
10	111	20.0	71	2	US-08-257-073-130	Sequence 130, App
11	97	17.5	2130	3	US-09-248-796A-4942	Sequence 4942, Ap
c 12	83.5	15.1	73519	3	US-09-949-016-16344	Sequence 16344, A
c 13	83.5	15.1	105919	3	US-09-949-016-11769	Sequence 11769, A
14	81.5	14.7	696	3	US-09-461-697-193	Sequence 193, App
15	81.5	14.7	699	3	US-09-461-697-191	Sequence 191, App
16	81.5	14.7	717	3	US-09-461-697-189	Sequence 189, App
17	81.5	14.7	774	3	US-09-461-697-187	Sequence 187, App
18	81.5	14.7	819	3	US-09-461-697-185	Sequence 185, App
19	81.5	14.7	1669	3	US-09-461-697-184	Sequence 184, App
20	80.5	14.5	1353	3	US-08-956-171E-521	Sequence 521, App
21	80.5	14.5	1353	3	US-08-781-986A-521	Sequence 521, App
22	80	14.4	1685	3	US-10-104-047-1957	Sequence 1957, Ap
23	79.5	14.4	5874	3	US-09-561-709B-2	Sequence 2, Appli
24	79	14.3	1149	3	US-09-252-991A-7034	Sequence 7034, Ap
25	79	14.3	1617	3	US-09-252-991A-6980	Sequence 6980, Ap
c 26	79	14.3	2751	3	US-09-252-991A-7066	Sequence 7066, Ap
27	78	14.1	1344	3	US-09-710-279-2899	Sequence 2899, Ap
28	78	14.1	1353	3	US-09-710-279-3225	Sequence 3225, Ap
c 29	78	14.1	3001	3	US-09-710-279-3813	Sequence 3813, Ap
30	78	14.1	3974	3	US-09-710-279-4245	Sequence 4245, Ap
c 31	78	14.1	28958	2	US-08-258-261B-6	Sequence 6, Appli
c 32	78	14.1	28958	2	US-08-456-837-6	Sequence 6, Appli
c 33	78	14.1	28958	2	US-08-457-342-6	Sequence 6, Appli
c 34	78	14.1	28958	2	US-08-457-646A-6	Sequence 6, Appli
c 35	78	14.1	28958	2	US-08-458-076A-6	Sequence 6, Appli
c 36	78	14.1	28958	2	US-08-764-233A-4	Sequence 4, Appli
c 37	78	14.1	28958	2	US-08-457-335A-6	Sequence 6, Appli
c 38	78	14.1	28958	2	US-08-729-214-6	Sequence 6, Appli
c 39	78	14.1	28958	3	US-09-028-934-6	Sequence 6, Appli
c 40	78	14.1	49377	2	US-08-764-233A-1	Sequence 1, Appli
41	77.5	14.0	1425	3	US-09-543-681A-2580	Sequence 2580, Ap
c 42	77	13.9	4413	3	US-10-104-047-514	Sequence 514, App
43	77	13.9	73519	3	US-09-949-016-16344	Sequence 16344, A
44	77	13.9	105919	3	US-09-949-016-11769	Sequence 11769, A
45	76.5	13.8	1243	3	US-09-270-767-14507	Sequence 14507, A

ALIGNMENTS

RESULT 1
US-08-098-327E-32
; Sequence 32, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; APPLICANT: DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/098,327E
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300

```
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-32

Alignment Scores:
Pred. No.:              7.01e-62        Length:         950
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-08-098-327E-32 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 2
US-08-462-625-32
; Sequence 32, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-32

Alignment Scores:
Pred. No.:              7.01e-62        Length:         950
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-08-462-625-32 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 3
US-08-098-327E-41
; Sequence 41, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
```

```
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-41

Alignment Scores:
Pred. No.:              1.3e-61        Length:         1482
Score:                  554.00         Matches:        107
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-837-344-19 (1-107) x US-08-098-327E-41 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 4
US-08-462-625-41
; Sequence 41, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-41

Alignment Scores:
Pred. No.:              1.3e-61        Length:         1482
Score:                  554.00         Matches:        107
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-837-344-19 (1-107) x US-08-462-625-41 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 5
US-08-098-327E-38
; Sequence 38, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-38

Alignment Scores:
Pred. No.:              1.31e-61        Length:         1493
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-08-098-327E-38 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 6
US-08-462-625-38
; Sequence 38, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  WO 92/13884
;       PUBLICATION DATE:  20-AUG-1992
US-08-462-625-38

Alignment Scores:
Pred. No.:              1.31e-61        Length:         1493
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-08-462-625-38 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 7
US-08-257-073-15
; Sequence 15, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-15

Alignment Scores:
Pred. No.:              1.49e-61        Length:         1320
Score:                  553.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-19 (1-107) x US-08-257-073-15 (1-1320)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        478 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 537

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 597

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 657

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 717

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 777

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        778 GAACAAAAAGAAGATAAATCT 798

RESULT 8
US-08-098-327E-45
; Sequence 45, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-45

Alignment Scores:
Pred. No.:              3.16e-61      Length:         1482
Score:                  551.00        Matches:        106
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  99.07%        Mismatches:     0
Query Match:            99.46%        Indels:         0
DB:                     3             Gaps:           0

US-09-837-344-19 (1-107) x US-08-098-327E-45 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 9
US-08-462-625-45
; Sequence 45, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-45

Alignment Scores:
Pred. No.:              3.16e-61      Length:         1482
Score:                  551.00        Matches:        106
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  99.07%        Mismatches:     0
Query Match:            99.46%        Indels:         0
DB:                     3             Gaps:           0

US-09-837-344-19 (1-107) x US-08-462-625-45 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960
```

RESULT 10
US-08-257-073-130
; Sequence 130, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-130

Alignment Scores:

Pred. No.:	2.14e-06	Length:	71
Score:	111.00	Matches:	23
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	20.04%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-19 (1-107) x US-08-257-073-130 (1-71)

```
Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          1 AGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 60

Qy         79 ValGluGly 81
              |||||||||
Db         61 GTTGAAGGA 69
```

RESULT 11
US-09-248-796A-4942
; Sequence 4942, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4942
; LENGTH: 2130
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4942

Alignment Scores:

Pred. No.:	0.0151	Length:	2130
Score:	97.00	Matches:	37
Percent Similarity:	44.00%	Conservative:	29
Best Local Similarity:	24.67%	Mismatches:	36
Query Match:	17.51%	Indels:	48
DB:	3	Gaps:	9

US-09-837-344-19 (1-107) x US-09-248-796A-4942 (1-2130)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIle------ 18
              ::::::   ||||||      |||::::::   |||||||||         |||
Db       1384 AAAGAAGTAGATACTGATTCTAATATTCAAGAAAAGAAGGAATTATTCCAAATTAAAGAG 1443

Qy         19 --------------------LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 31
                                  :::   |||||||||      :::::::|||   |||:::
Db       1444 CCAGTAAAAGATGAAAAACAAATTAAAGAAGACTTGGTC---AAAGTTGAAACCCCATCC 1500

Qy         32 --------------------IleGluLeuProSerGluAsnGluArgGlyTyrTyrIle 44
                                  :::|||:::||||||::::::|||      :::   |||
Db       1501 AAAGATGAAGGACCGGCGAAGGTTGAAGTTCCATCCAAAGATGAAACACCATTCAAGATT 1560

Qy         45 ProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLys------ 62
                 :::      |||         ||||||   |||   ::::::|||   |||
Db       1561 CAAAATATTGATAGT---ACCACTGATAAT---GGCAAAGCTAAAGATGAGAAAGATAAT 1614

Qy         63 ------------------------GluIleSerIleIleGluLysThrAsnArgGluSer 74
                                     :::     ::::::   |||   :::::::::
Db       1615 GTTTCTAAAGTTGATTCTAACGTTGATACTAAAGTAGTTACCAAAGACGATCAAGATGGT 1674

Qy         75 Ile----------------------------ThrThrAsnValGluGlyArgArgAsp 84
              |||                          :::   ||||||::::::|||
Db       1675 ATTAAAAAAGTAGAAGAGAAGCCTGTTGAAAAGAAAGACGACGACGAAGGTAAGAAAGAC 1734

Qy         85 --------------IleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLys 99
                           :::|||||   :::   ||||||   ||||||      :::
Db       1735 AATGAAAAGGTTACTGTTCACAAGGACAATGAGGAAGAAGATAAGGATAAGGGTGTCAAT 1794

Qy        100 ProGlu------------GlnLysGluAsp 105
                             ||||||:::|||
Db       1795 GAAGCCAAAGATAAGGTTCAGAAAGATGAT 1824
```

RESULT 12
US-09-949-016-16344/c
; Sequence 16344, Application US/09949016

```
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16344
;   LENGTH: 73519
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(73519)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16344

Alignment Scores:
Pred. No.:              111          Length:        73519
Score:                  83.50        Matches:       37
Percent Similarity:     45.08%       Conservative:  18
Best Local Similarity:  30.33%       Mismatches:    28
Query Match:            15.07%       Indels:        39
DB:                     3            Gaps:          9

US-09-837-344-19 (1-107) x US-09-949-016-16344 (1-73519)

Qy          6 LysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr--- 24
              |||      ||||||:::|||:::      |||          |||          |||
Db      61539 AAGACAAGATTAGAGAAAAAAGAATGAAAAGGAACAAACAAAGCCTCCAAGAAATACAGG 61480

Qy         25 -------------------------------------------------GlyArgLeu 27
                                                                 ||||||:::
Db      61479 ACTCTGTGAAAAGACCAAACCTACGTTCAGTTGGTGTACCTGAAAGTGACGGGGAGAATG 61420

Qy         28 GluIleProAlaIleGluLeuProSer---GluAsnGluArgGlyTyrTyr--------- 43
              |||                 ||||||   ::::::      |||||||||
Db      61419 GAA-----------------CCAAGTTGAAAACACTCTTCGGGGTATTATCCAGGAGAA 61378

Qy         44 IleProHisGlnSerSerLeuProGln---AspAsnArgGlyAsnSerArgAspSerLys 62
              :::|||      |||          |||      |||   ||||||:::   ::::::
Db      61377 CTTCCCCAACCTAGCAAGACAGGCCAATATTCAAATTCGGGAAATACAGAGAACACCACA 61318

Qy         63 GluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArg 82
              :::|||   ::::::|||||||||||||||:::      |||   |||   :::
Db      61317 AAGATA---CTCCTCGAGAAGACCAACCGCAAG------ACACATAATTGTCAG------ 61273

Qy         83 ArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGln 102
                 ||||||||||||      |||   |||         :::::::::      |||
Db      61272 ------ATTCACAAAGGTTGAAATGAAGGAAAA---------AATGTTAGGGGCAGCCAG 61228

Qy        103 LysGlu 104
              :::|||
Db      61227 AGGGAA 61222

RESULT 13
US-09-949-016-11769/c
; Sequence 11769, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11769
;   LENGTH: 105919
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(105919)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11769

Alignment Scores:
Pred. No.:              184          Length:        105919
Score:                  83.50        Matches:       37
Percent Similarity:     45.08%       Conservative:  18
Best Local Similarity:  30.33%       Mismatches:    28
Query Match:            15.07%       Indels:        39
DB:                     3            Gaps:          9

US-09-837-344-19 (1-107) x US-09-949-016-11769 (1-105919)

Qy          6 LysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr--- 24
              |||      ||||||:::|||:::      |||          |||          |||
Db      93940 AAGACAAGATTAGAGAAAAAAGAATGAAAAGGAACAAACAAAGCCTCCAAGAAATACAGG 93881

Qy         25 -------------------------------------------------GlyArgLeu 27
                                                                 ||||||:::
Db      93880 ACTCTGTGAAAAGACCAAACCTACGTTCAGTTGGTGTACCTGAAAGTGACGGGGAGAATG 93821

Qy         28 GluIleProAlaIleGluLeuProSer---GluAsnGluArgGlyTyrTyr--------- 43
              |||                 ||||||   ::::::      |||||||||
Db      93820 GAA-----------------CCAAGTTGAAAACACTCTTCGGGGTATTATCCAGGAGAA 93779

Qy         44 IleProHisGlnSerSerLeuProGln---AspAsnArgGlyAsnSerArgAspSerLys 62
              :::|||      |||          |||      |||   ||||||:::   ::::::
Db      93778 CTTCCCCAACCTAGCAAGACAGGCCAATATTCAAATTCGGGAAATACAGAGAACACCACA 93719

Qy         63 GluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArg 82
              :::|||   ::::::|||||||||||||||:::      |||   |||   :::
Db      93718 AAGATA---CTCCTCGAGAAGACCAACCGCAAG------ACACATAATTGTCAG------ 93674

Qy         83 ArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGln 102
                 ||||||||||||      |||   |||         :::::::::      |||
Db      93673 ------ATTCACAAAGGTTGAAATGAAGGAAAA---------AATGTTAGGGGCAGCCAG 93629

Qy        103 LysGlu 104
              :::|||
Db      93628 AGGGAA 93623

RESULT 14
US-09-461-697-193
; Sequence 193, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
```

```
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 193
;   LENGTH: 696
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-193

Alignment Scores:
Pred. No.:              0.318          Length:         696
Score:                  81.50          Matches:        31
Percent Similarity:     44.34%         Conservative:   16
Best Local Similarity:  29.25%         Mismatches:     54
Query Match:            14.71%         Indels:         5
DB:                     3              Gaps:           3

US-09-837-344-19 (1-107) x US-09-461-697-193 (1-696)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        232 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 291

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||       :::   ||||||:::|||
Db        292 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 351

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||       |||
Db        352 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 411

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||             ||||||         :::         |||   |||
Db        412 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 471

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||       :::
Db        472 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 531

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        532 AAAGAAGATGAAAAAGAG 549

RESULT 15
US-09-461-697-191
; Sequence 191, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 191
;   LENGTH: 699
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-191

Alignment Scores:
Pred. No.:              0.319          Length:         699
Score:                  81.50          Matches:        31
Percent Similarity:     44.34%         Conservative:   16
Best Local Similarity:  29.25%         Mismatches:     54
Query Match:            14.71%         Indels:         5
DB:                     3              Gaps:           3

US-09-837-344-19 (1-107) x US-09-461-697-191 (1-699)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        235 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 294

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||       :::   ||||||:::|||
Db        295 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 354

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||       |||
Db        355 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 414

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||             ||||||         :::         |||   |||
Db        415 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 474

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||       :::
Db        475 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 534

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        535 AAAGAAGATGAAAAAGAG 552

RESULT 16
US-09-461-697-189
; Sequence 189, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 189
;   LENGTH: 717
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-189

Alignment Scores:
Pred. No.:              0.331          Length:         717
Score:                  81.50          Matches:        31
Percent Similarity:     44.34%         Conservative:   16
Best Local Similarity:  29.25%         Mismatches:     54
Query Match:            14.71%         Indels:         5
DB:                     3              Gaps:           3

US-09-837-344-19 (1-107) x US-09-461-697-189 (1-717)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        253 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 312

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
```

```
                    :::   |||            |||      :::   ||||||:::|||
Db        313 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 372

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                       :::                  :::|||   :::|||||      |||
Db        373 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 432

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::          |||   |||
Db        433 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 492

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||            ||||||      :::|||   ||||||      :::
Db        493 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 552

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        553 AAAGAAGATGAAAAAGAG 570

RESULT 17
US-09-461-697-187
; Sequence 187, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187
;   LENGTH: 774
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-187

Alignment Scores:
Pred. No.:              0.368          Length:         774
Score:                  81.50          Matches:        31
Percent Similarity:     44.34%         Conservative:   16
Best Local Similarity:  29.25%         Mismatches:     54
Query Match:            14.71%         Indels:         5
DB:                     3              Gaps:           3

US-09-837-344-19 (1-107) x US-09-461-697-187 (1-774)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        310 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 369

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                    :::   |||            |||      :::   ||||||:::|||
Db        370 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 429

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                       :::                  :::|||   :::|||||      |||
Db        430 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 489

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::          |||   |||
Db        490 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 549

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||            ||||||      :::|||   ||||||      :::
Db        550 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 609

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        610 AAAGAAGATGAAAAAGAG 627

RESULT 18
US-09-461-697-185
; Sequence 185, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 185
;   LENGTH: 819
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-185

Alignment Scores:
Pred. No.:              0.398          Length:         819
Score:                  81.50          Matches:        31
Percent Similarity:     44.34%         Conservative:   16
Best Local Similarity:  29.25%         Mismatches:     54
Query Match:            14.71%         Indels:         5
DB:                     3              Gaps:           3

US-09-837-344-19 (1-107) x US-09-461-697-185 (1-819)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        355 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 414

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                    :::   |||            |||      :::   ||||||:::|||
Db        415 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 474

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                       :::                  :::|||   :::|||||      |||
Db        475 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 534

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::          |||   |||
Db        535 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 594

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||            ||||||      :::|||   ||||||      :::
Db        595 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 654

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        655 AAAGAAGATGAAAAAGAG 672

RESULT 19
US-09-461-697-184
; Sequence 184, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
```

```
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184
;   LENGTH: 1669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-461-697-184

Alignment Scores:
Pred. No.:              1.07          Length:          1669
Score:                  81.50         Matches:         31
Percent Similarity:     44.34%        Conservative:    16
Best Local Similarity:  29.25%        Mismatches:      54
Query Match:            14.71%        Indels:          5
DB:                     3             Gaps:            3

US-09-837-344-19 (1-107) x US-09-461-697-184 (1-1669)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::          ||||||
Db        434 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 493

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||              |||       :::   ||||||:::|||
Db        494 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 553

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                  :::|||   :::|||||||      |||
Db        554 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 613

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
                 ||||||              ||||||         :::          |||   |||
Db        614 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 673

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                ||||||     :::|||   ||||||     :::
Db        674 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 733

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::|||||||
Db        734 AAAGAAGATGAAAAAGAG 751

RESULT 20
US-08-956-171E-521
; Sequence 521, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 521:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1353 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 521:
US-08-956-171E-521

Alignment Scores:
Pred. No.:              1.07          Length:          1353
Score:                  80.50         Matches:         30
Percent Similarity:     45.38%        Conservative:    24
Best Local Similarity:  25.21%        Mismatches:      39
Query Match:            14.53%        Indels:          26
DB:                     3             Gaps:            4

US-09-837-344-19 (1-107) x US-08-956-171E-521 (1-1353)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        461 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 520

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::      ::::::                   ||||||   :::|||
Db        521 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 580

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                    :::   |||:::   :::            |||                 :::
Db        581 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 640

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                  ||| |||:::      :::|||:::   ||||||:::   |||:::::
Db        641 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 700

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                                    :::|||   |||
Db        701 CACAAAC----------------------------------------CTGCATCTCAAGAA 721

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 |||      ||||||          ||||||   ::::::||||||   :::
Db        722 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 776

RESULT 21
US-08-781-986A-521
; Sequence 521, Application US/08781986A
; Patent No. 6737248
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  521:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1353 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-521

Alignment Scores:
Pred. No.:                1.07          Length:        1353
Score:                    80.50         Matches:       30
Percent Similarity:       45.38%        Conservative:  24
Best Local Similarity:    25.21%        Mismatches:    39
Query Match:              14.53%        Indels:        26
DB:                       3             Gaps:          4

US-09-837-344-19 (1-107) x US-08-781-986A-521 (1-1353)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        461 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 520

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::     ::::::                 ||||||   :::|||
Db        521 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 580

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::            |||                 :::
Db        581 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 640

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                  ||| |||:::     :::|||:::   ||||||:::   |||::::::
Db        641 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 700

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                                        :::|||   |||
Db        701 CACAAAC-----------------------------------CTGCATCTCAAGAA 721

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
              |||      ||||||         ||||||   ::::::||||||   :::
Db        722 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 776

RESULT 22
US-10-104-047-1957
; Sequence 1957, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. 6943241el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1957
;   LENGTH: 1685
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-104-047-1957

Alignment Scores:
Pred. No.:                1.68          Length:        1685
Score:                    80.00         Matches:       29
Percent Similarity:       43.24%        Conservative:  19
Best Local Similarity:    26.13%        Mismatches:    53
Query Match:              14.44%        Indels:        10
DB:                       3             Gaps:          3

US-09-837-344-19 (1-107) x US-10-104-047-1957 (1-1685)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||:::         :::         ::::::   :::|||   |||:::
Db        707 AGAAGAAGATCAAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAA 766

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||      :::   |||           |||      :::   ||||||:::
Db        767 GAAGACAAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAA 826

Qy         41 GlyTyrTyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSer 58
              |||                :::                 :::|||   :::|||||||
Db        827 GGAGAAGATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGA 886

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr--- 77
                 |||   ||||||              ||||||         :::          |||
Db        887 GAAGATGGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGG 946

Qy         78 ------------AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLys 93
                         |||   :::|||:::             ||||||      :::|||
Db        947 GATGGAAAAGAGAATGAAGATGGAAAA---------GAGAAGGGAGATAAAAAAGAGGGG 997

Qy         94 LysAspGlySerIleLysProGluGlnLysGlu 104
              ||||||      :::|||   ::::::||||||
Db        998 AAAGATGTAAAAGTCAAAGAAGATGAAAAAGAG 1030

RESULT 23
US-09-561-709B-2
; Sequence 2, Application US/09561709B
; Patent No. 6682911
; GENERAL INFORMATION:
;  APPLICANT: Burgeson, Robert
;  APPLICANT:  Champliaud, Marie-France
;  APPLICANT:  Olson, Pamela
;  APPLICANT:  Koch, Manuel
;  APPLICANT:  Brunken, William
;  TITLE OF INVENTION: LAMININS AND USES THEREOF
;  FILE REFERENCE: 10287-060001
;  CURRENT APPLICATION NUMBER: US/09/561,709B
;  CURRENT FILING DATE:  2000-05-01
;  PRIOR APPLICATION NUMBER: US 09/168,949
;  PRIOR FILING DATE: 1998-10-09
;  PRIOR APPLICATION NUMBER: US 60/061,609
;  PRIOR FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 5874
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (81)...(5363)
US-09-561-709B-2

Alignment Scores:
Pred. No.:                11              Length:         5874
Score:                    79.50           Matches:        32
Percent Similarity:       42.73%          Conservative:   15
Best Local Similarity:    29.09%          Mismatches:     34
Query Match:              14.35%          Indels:         29
DB:                       3               Gaps:           5

US-09-837-344-19 (1-107) x US-09-561-709B-2 (1-5874)

Qy          4 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeu 23
              |||   |||      :::||||||   |||         |||:::|||   |||||||||
Db       3888 GATCTGAAAGATACAATAGAAAGAGCAAAGAATGAAGCAGACCTCTTACTTGAAGACCTT 3947

Qy         24 TyrGlyArgLeuGluIleProAlaIleGluLeu-----------------ProSerGlu 37
                        :::::::::   :::      |||                  ||||||
Db       3948 CAGGAAGAAATTGATTTGCAATCCAGTGTCCTTAATGCAAGCATTGCGGACTCCTCAGAA 4007

Qy         38 AsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsn 57
              |||   :::   ||||||      |||   ||||||
Db       4008 AACATCAAGAAATATTAT------CACATATCATCA----------------------- 4037

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
                    |||   |||       |||   |||   ::::::      :::|||   |||
Db       4038 ---------TCTGCTGAAAAGAAAATTAATGAAACTAGTTCC------ACCATTAATACC 4082

Qy         78 AsnValGluGlyArgArgAspIle----------------------HisLysGlyHis 89
              :::         |||   |||:::                     ||||||:::
Db       4083 TCTGCAAATACAAGGAATGACTTACTTACCATCTTAGATACACTAACCTCAAAAGGAAAC 4142

Qy         90 LeuGluGluLysLysAspGlySerIleLys 99
              |||       ::::::       ||||||
Db       4143 TTGTCATTGGAAAGATTAAAGCAGATTAAG 4172

RESULT 24
US-09-252-991A-7034
; Sequence 7034, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 7034
;   LENGTH: 1149
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7034

Alignment Scores:
Pred. No.:                1.33            Length:         1149
Score:                    79.00           Matches:        34
Percent Similarity:       39.37%          Conservative:   16
Best Local Similarity:    26.77%          Mismatches:     41
Query Match:              14.26%          Indels:         36
DB:                       3               Gaps:           5

US-09-837-344-19 (1-107) x US-09-252-991A-7034 (1-1149)

Qy          5 ThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr 24
              |||   :::|||||||||:::          |||          |||||||||:::
Db        446 ACCCACCGCAACCTGGAGAAGATGGTCGCCGAGGGCAGC---TTCCGCGAGGACCTGTTC 502

Qy         25 GlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIle 44
                ||||||   :::   ::::::|||:::|||          |||||||||:::   |||
Db        503 TACCGCCTCAACGTGCTCAGCCTGGAAGTCCCGCCGTTGCGAGAGCGCGGCCACGACATC 562

Qy         45 ------------ProHis------GlnSerSerLeuProGlnAspAsnArg--------- 55
                          |||       |||   :::       |||       |||
Db        563 CTGCTGCTGGCGCGGCACTTCATGCAGCAGGCCTGTGCGCAGATCCAGCGCCCGGTGTGC 622

Qy         56 -------------------------------------------------GlyAsnSerArg 59
                                                              ||||||   |||
Db        623 CGGCTGGCGCCGGGCACCTACCCGGCGCTGCTGAGCAACCGCTGGCCGGGCAACGTGCGC 682

Qy         60 AspSerLys------------------------------GluIleSerIleIleGluLys 69
                  :::                              |||   |||:::::::::
Db        683 CAGTTGCAGAACGTGATCTTCCGTGCCGCTGCGATCTGCGAGAGCAGCCTGGTGGATATC 742

Qy         70 ThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHis 89
              :::   |||          |||   |||      :::   |||          |||
Db        743 GGCGACCTGGAGATCGCCGGCACCGCCGTGGCGCGGCAGAACGACGGCGAGGTGGGCAGC 802

Qy         90 LeuGluGluLysLysAspGly 96
              |||||||||       :::|||
Db        803 CTGGAGGAGGCGGTGGAGGGC 823

RESULT 25
US-09-252-991A-6980
; Sequence 6980, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6980
;   LENGTH: 1617
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6980

Alignment Scores:
Pred. No.:                2.14            Length:         1617
Score:                    79.00           Matches:        34
Percent Similarity:       39.37%          Conservative:   16
Best Local Similarity:    26.77%          Mismatches:     41
Query Match:              14.26%          Indels:         36
DB:                       3               Gaps:           5

US-09-837-344-19 (1-107) x US-09-252-991A-6980 (1-1617)

Qy          5 ThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr 24
              |||   :::|||||||||:::          |||          |||||||||:::
Db       1108 ACCCACCGCAACCTGGAGAAGATGGTCGCCGAGGGCAGC---TTCCGCGAGGACCTGTTC 1164

Qy         25 GlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIle 44
                ||||||   :::   ::::::|||:::|||          |||||||||:::   |||
Db       1165 TACCGCCTCAACGTGCTCAGCCTGGAAGTCCCGCCGTTGCGAGAGCGCGGCCACGACATC 1224

Qy         45 ------------ProHis------GlnSerSerLeuProGlnAspAsnArg--------- 55
                          |||       |||   :::       |||       |||
Db       1225 CTGCTGCTGGCGCGGCACTTCATGCAGCAGGCCTGTGCGCAGATCCAGCGCCCGGTGTGC 1284

Qy         56 -------------------------------------------------GlyAsnSerArg 59
                                                              ||||||   |||
Db       1285 CGGCTGGCGCCGGGCACCTACCCGGCGCTGCTGAGCAACCGCTGGCCGGGCAACGTGCGC 1344
```

```
Qy          60 AspSerLys----------------------------GluIleSerIleIleGluLys 69
               :::                                  |||   |||:::::::::
Db        1345 CAGTTGCAGAACGTGATCTTCCGTGCCGCTGCGATCTGCGAGAGCAGCCTGGTGGATATC 1404

Qy          70 ThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHis 89
                  :::   |||         |||   |||      :::  |||         |||
Db        1405 GGCGACCTGGAGATCGCCGGCACCGCCGTGGCGCGGCAGAACGACGGCGAGGTGGGCAGC 1464

Qy          90 LeuGluGluLysLysAspGly 96
               |||||||||      :::|||
Db        1465 CTGGAGGAGGCGGTGGAGGGC 1485

RESULT 26
US-09-252-991A-7066/c
; Sequence 7066, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 7066
;   LENGTH: 2751
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7066

Alignment Scores:
Pred. No.:              4.47          Length:         2751
Score:                  79.00         Matches:        34
Percent Similarity:     39.37%        Conservative:   16
Best Local Similarity:  26.77%        Mismatches:     41
Query Match:            14.26%        Indels:         36
DB:                     3             Gaps:           5

US-09-837-344-19 (1-107) x US-09-252-991A-7066 (1-2751)

Qy           5 ThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr 24
               |||   :::|||||||||:::         |||             |||||||||:::
Db         561 ACCCACCGCAACCTGGAGAAGATGGTCGCCGAGGGCAGC---TTCCGCGAGGACCTGTTC 505

Qy          25 GlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIle 44
               ||||||   :::   ::::::|||:::|||          |||||||||:::   |||
Db         504 TACCGCCTCAACGTGCTCAGCCTGGAAGTCCCGCCGTTGCGAGAGCGCGGCCACGACATC 445

Qy          45 ------------ProHis------GlnSerSerLeuProGlnAspAsnArg--------- 55
                         |||      |||   :::      |||      |||
Db         444 CTGCTGCTGGCGCGGCACTTCATGCAGCAGGCCTGTGCGCAGATCCAGCGCCCGGTGTGC 385

Qy          56 -------------------------------------------------GlyAsnSerArg 59
                                                              ||||||   |||
Db         384 CGGCTGGCGCCGGGCACCTACCCGGCGCTGCTGAGCAACCGCTGGCCGGGCAACGTGCGC 325

Qy          60 AspSerLys----------------------------GluIleSerIleIleGluLys 69
                  :::                               |||   |||:::::::::
Db         324 CAGTTGCAGAACGTGATCTTCCGTGCCGCTGCGATCTGCGAGAGCAGCCTGGTGGATATC 265

Qy          70 ThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHis 89
                  :::   |||         |||   |||      :::  |||         |||
Db         264 GGCGACCTGGAGATCGCCGGCACCGCCGTGGCGCGGCAGAACGACGGCGAGGTGGGCAGC 205

Qy          90 LeuGluGluLysLysAspGly 96
               |||||||||      :::|||
Db         204 CTGGAGGAGGCGGTGGAGGGC 184

RESULT 27
US-09-710-279-2899
; Sequence 2899, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2899
;   LENGTH: 1344
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-2899

Alignment Scores:
Pred. No.:              2.23          Length:         1344
Score:                  78.00         Matches:        27
Percent Similarity:     43.12%        Conservative:   20
Best Local Similarity:  24.77%        Mismatches:     36
Query Match:            14.08%        Indels:         26
DB:                     3             Gaps:           4

US-09-837-344-19 (1-107) x US-09-710-279-2899 (1-1344)

Qy           4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
               :::|||   :::||||||:::               :::|||||||||:::   ::::::
Db          82 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 141

Qy          19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                  |||   |||          :::|||       |||
Db         142 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA---------------------- 174

Qy          39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                              |||   |||          |||:::|||
Db         175 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 216

Qy          59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
                  |||:::|||:::   :::      |||::::::      :::
Db         217 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 276

Qy          79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                  ||||||       |||                  |||||||||   |||   :::
Db         277 CCAGAAGGAACTTCAGAT-----------------GAAGAAAAACCTGATGATGCACCA 318

Qy          99 LysProGluGlnLysGluAspLysSer 107
               ||||||:::   |||   :::::::::
Db         319 AAACCAGATGACAAACCTGAAGAAACA 345

RESULT 28
US-09-710-279-3225
; Sequence 3225, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3225
;   LENGTH: 1353
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3225

Alignment Scores:
Pred. No.:                2.25          Length:          1353
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       3             Gaps:            4

US-09-837-344-19 (1-107) x US-09-710-279-3225 (1-1353)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db         82 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 141

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db        142 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 174

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                              |||   |||         |||:::|||
Db        175 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 216

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              .|||:::|||:::   :::      |||::::::     :::
Db        217 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 276

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              .||||||      |||                 |||||||||   |||   :::
Db        277 CCAGAAGGAACTTCAGAT-----------------GAAGAAAAACCTGATGATGCACCA 318

Qy         99 LysProGluGlnLysGluAspLysSer 107
              ||||||:::   |||   :::::::::
Db        319 AAACCAGATGACAAACCTGAAGAAACA 345

RESULT 29
US-09-710-279-3813/c
; Sequence 3813, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3813
;   LENGTH: 3001
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3813

Alignment Scores:
Pred. No.:                6.78          Length:          3001
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       3             Gaps:            4

US-09-837-344-19 (1-107) x US-09-710-279-3813 (1-3001)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db       2417 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 2358

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db       2357 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 2325

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                              |||   |||         |||:::|||
Db       2324 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 2283

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              .|||:::|||:::   :::      |||::::::     :::
Db       2282 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 2223

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              ||||||      |||                 |||||||||   |||   :::
Db       2222 CCAGAAGGAACTTCAGAT-----------------GAAGAAAAACCTGATGATGCACCA 2181

Qy         99 LysProGluGlnLysGluAspLysSer 107
              ||||||:::   |||   :::::::::
Db       2180 AAACCAGATGACAAACCTGAAGAAACA 2154

RESULT 30
US-09-710-279-4245
; Sequence 4245, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4245
;   LENGTH: 3974
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4245

Alignment Scores:
Pred. No.:                9.99          Length:          3974
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       3             Gaps:            4

US-09-837-344-19 (1-107) x US-09-710-279-4245 (1-3974)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db        499 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 558

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db        559 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 591

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                              |||   |||         |||:::|||
Db        592 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 633

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              .|||:::|||:::   :::      |||::::::     :::
Db        634 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 693
```

```
Qy          79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
               ||||||      |||                  |||||||||   |||   :::
Db         694 CCAGAAGGAACTTCAGAT----------------GAAGAAAAACCTGATGATGCACCA 735

Qy          99 LysProGluGlnLysGluAspLysSer 107
               ||||||:::   |||   :::::::::
Db         736 AAACCAGATGACAAACCTGAAGAAACA 762


Search completed: December 20, 2005, 05:12:41
Job time : 203 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 00:19:33 ; Search time 691 Seconds
(without alignments)
1280.497 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103709_293/app_query.fasta_1.263
-DB=Published_Applications_NA_Main -QFMT=fastap -SUFFIX=rnpbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_1026_@runat_15122005_103709_293
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	950	3	US-09-837-344-32	Sequence 32, Appl
2	554	100.0	950	3	US-09-900-963-32	Sequence 32, Appl
3	554	100.0	1482	3	US-09-837-344-41	Sequence 41, Appl
4	554	100.0	1482	3	US-09-900-963-41	Sequence 41, Appl
5	554	100.0	1493	3	US-09-837-344-38	Sequence 38, Appl
6	554	100.0	1493	3	US-09-900-963-38	Sequence 38, Appl
7	551	99.5	1482	3	US-09-837-344-45	Sequence 45, Appl
8	551	99.5	1482	3	US-09-900-963-45	Sequence 45, Appl
9	94	17.0	1385	3	US-09-894-018-122	Sequence 122, App
10	94	17.0	1385	8	US-10-474-960A-122	Sequence 122, App
11	83.5	15.1	630	4	US-09-925-065A-788059	Sequence 788059,
12	83.5	15.1	630	4	US-09-925-065A-846536	Sequence 846536,
13	81.5	14.7	696	3	US-09-922-261-193	Sequence 193, App
14	81.5	14.7	699	3	US-09-922-261-191	Sequence 191, App
15	81.5	14.7	717	3	US-09-922-261-189	Sequence 189, App
16	81.5	14.7	774	3	US-09-922-261-187	Sequence 187, App
17	81.5	14.7	819	3	US-09-922-261-185	Sequence 185, App
18	81.5	14.7	1455	3	US-09-769-736-71	Sequence 71, Appl
19	81.5	14.7	1669	3	US-09-922-261-184	Sequence 184, App
20	81.5	14.7	2460	9	US-10-450-763-10274	Sequence 10274, A
21	81	14.6	1508	7	US-10-425-114-6320	Sequence 6320, Ap
c 22	81	14.6	1689	7	US-10-424-599-91249	Sequence 91249, A
23	81	14.6	1701	7	US-10-424-599-25309	Sequence 25309, A
24	80.5	14.5	540	8	US-10-357-930-49495	Sequence 49495, A
25	80.5	14.5	630	4	US-09-925-065A-846535	Sequence 846535,
26	80.5	14.5	1353	2	US-08-781-986A-521	Sequence 521, App
27	80.5	14.5	1353	7	US-10-329-624-521	Sequence 521, App
28	80.5	14.5	1524	7	US-10-381-596A-3	Sequence 3, Appli
29	80.5	14.5	2709	7	US-10-282-122A-16144	Sequence 16144, A
30	80	14.4	1685	6	US-10-104-047-1957	Sequence 1957, Ap
31	80	14.4	2193	8	US-10-739-930-1417	Sequence 1417, Ap
c 32	80	14.4	3684	10	US-11-097-143-32615	Sequence 32615, A
33	79.5	14.4	700	5	US-10-040-739-1385	Sequence 1385, Ap
34	79.5	14.4	966	3	US-09-864-761-17222	Sequence 17222, A
35	79.5	14.4	1086	3	US-09-864-761-17120	Sequence 17120, A
36	79.5	14.4	1086	3	US-09-864-761-17617	Sequence 17617, A
37	79.5	14.4	3144	6	US-10-029-386-24671	Sequence 24671, A
38	79.5	14.4	3264	6	US-10-029-386-20776	Sequence 20776, A
39	79.5	14.4	4972	7	US-10-764-425-77	Sequence 77, Appl
40	79.5	14.4	6149	9	US-10-450-763-22369	Sequence 22369, A
41	79.5	14.4	6370	9	US-10-450-763-22370	Sequence 22370, A
42	78	14.1	1397	7	US-10-425-114-31980	Sequence 31980, A
43	78	14.1	1397	8	US-10-425-115-42134	Sequence 42134, A
44	78	14.1	1449	7	US-10-425-114-4694	Sequence 4694, Ap
45	78	14.1	1449	8	US-10-425-115-42131	Sequence 42131, A

ALIGNMENTS

RESULT 1
US-09-837-344-32
; Sequence 32, Application US/09837344
; Patent No. US20020041882A1
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/837,344
; FILING DATE: 19-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/462,625
; FILING DATE: <Unknown>
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991

```
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-837-344-32

Alignment Scores:
Pred. No.:              2.24e-63        Length:         950
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-837-344-32 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 2
US-09-900-963-32
; Sequence 32, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-900-963-32

Alignment Scores:
Pred. No.:              2.24e-63        Length:         950
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-900-963-32 (1-950)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 689

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 749

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 809

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 869

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 929

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        930 GAACAAAAAGAAGATAAATCT 950

RESULT 3
US-09-837-344-41
; Sequence 41, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
```

```
;                          HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                          OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/837,344
;            FILING DATE: 19-Apr-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/462,625
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: FR 91 01286
;            FILING DATE: 05-FEB-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-078
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 41:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1482 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1482
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-837-344-41

Alignment Scores:
Pred. No.:              4.04e-63        Length:         1482
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-837-344-41 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 4
US-09-900-963-41
; Sequence 41, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 41:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1482 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1482
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-900-963-41

Alignment Scores:
Pred. No.:              4.04e-63        Length:         1482
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-900-963-41 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 5
US-09-837-344-38
; Sequence 38, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1493 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-837-344-38

Alignment Scores:
Pred. No.:              4.08e-63      Length:          1493
Score:                  554.00        Matches:         107
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     3             Gaps:            0

US-09-837-344-19 (1-107) x US-09-837-344-38 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 6
US-09-900-963-38
; Sequence 38, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 38:
```

```
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1493 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-900-963-38

Alignment Scores:
Pred. No.:                4.08e-63        Length:          1493
Score:                    554.00          Matches:         107
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-837-344-19 (1-107) x US-09-900-963-38 (1-1493)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 696

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 756

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 816

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 876

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 936

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        937 GAACAAAAAGAAGATAAATCT 957

RESULT 7
US-09-837-344-45
; Sequence 45, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 45:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1482 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..1482
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-837-344-45

Alignment Scores:
Pred. No.:                1.01e-62        Length:          1482
Score:                    551.00          Matches:         106
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    99.07%          Mismatches:      0
Query Match:              99.46%          Indels:          0
DB:                       3               Gaps:            0

US-09-837-344-19 (1-107) x US-09-837-344-45 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 8
US-09-900-963-45
; Sequence 45, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
```

```
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 45:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1482 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1482
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-900-963-45

Alignment Scores:
Pred. No.:              1.01e-62        Length:         1482
Score:                  551.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.46%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-900-963-45 (1-1482)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Qy        101 GluGlnLysGluAspLysSer 107
              |||||||||||||||||||||
Db        940 GAACAAAAAGAAGATAAATCT 960

RESULT 9
US-09-894-018-122
; Sequence 122, Application US/09894018
; Patent No. US20020119127A1
; GENERAL INFORMATION:
;  APPLICANT: EPIMMUNE, Inc.
;  APPLICANT:  Sette, Alessandro
;  APPLICANT:  Chestnut, Robert
;  APPLICANT:  Livingston, Brian
;  APPLICANT:  Baker, Denisw
;  APPLICANT:  Newman, Mark
;  APPLICANT:  Brown, David
;  TITLE OF INVENTION: METODS AND SYSTEM FOR OPTIMIZING
;  TITLE OF INVENTION:  MINIGENES AND PEPTIDES THEREBY
;  FILE REFERENCE: 39963-20033.00
;  CURRENT APPLICATION NUMBER: US/09/894,018
;  CURRENT FILING DATE:  2001-06-27
;  PRIOR APPLICATION NUMBER: PCT/US00/35568
;  PRIOR FILING DATE: 2000-12-28
;  PRIOR APPLICATION NUMBER: US 60/173,390
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: US 60/284,221
;  PRIOR FILING DATE: 2001-04-16
;  NUMBER OF SEQ ID NOS: 368
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 122
;   LENGTH: 1385
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Encoding sequence for PfCTL/HTL(N)
US-09-894-018-122

Alignment Scores:
Pred. No.:              0.0548          Length:         1385
Score:                  94.00           Matches:        18
Percent Similarity:     85.71%          Conservative:   0
Best Local Similarity:  85.71%          Mismatches:     3
Query Match:            16.97%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-19 (1-107) x US-09-894-018-122 (1-1385)

Qy         30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
              |||          ||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTGGACCAGGTCTGCCGTCCGAAAACGAAAGAGGATACTACATACCTCACCAGAGCAGC 420

Qy         50 Leu 50
              |||
Db        421 CTC 423

RESULT 10
US-10-474-960A-122
; Sequence 122, Application US/10474960A
; Publication No. US20040248113A1
; GENERAL INFORMATION:
;  APPLICANT: Sette, Alessandro
;  APPLICANT:  Chesnut, Robert
;  APPLICANT:  Livingston, Brian
;  APPLICANT:  Baker, Denise
;  APPLICANT:  Newman, Mark
;  APPLICANT:  Brown, David
;  TITLE OF INVENTION: Methods and System for Optimizing Multi-epitope Nucleic
;  TITLE OF INVENTION:  Acid Constructs and Peptides Encoded Thereby
;  FILE REFERENCE: 2060.0320004
;  CURRENT APPLICATION NUMBER: US/10/474,960A
;  CURRENT FILING DATE:  2003-10-16
;  PRIOR APPLICATION NUMBER: PCT/US02/09877
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 09/894,018
```

```
;  PRIOR FILING DATE: 2001-06-27
;  PRIOR APPLICATION NUMBER: US 60/284,221
;  PRIOR FILING DATE: 2001-04-16
;  NUMBER OF SEQ ID NOS: 419
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 122
;   LENGTH: 1385
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Encoding sequence for PfCTL/HTL(N)
US-10-474-960A-122

Alignment Scores:
Pred. No.:              0.0548          Length:         1385
Score:                  94.00           Matches:        18
Percent Similarity:     85.71%          Conservative:   0
Best Local Similarity:  85.71%          Mismatches:     3
Query Match:            16.97%          Indels:         0
DB:                     8               Gaps:           0

US-09-837-344-19 (1-107) x US-10-474-960A-122 (1-1385)

Qy         30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
              |||          ||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTGGACCAGGTCTGCCGTCCGAAAACGAAAGAGGATACTACATACCTCACCAGAGCAGC 420

Qy         50 Leu 50
              |||
Db        421 CTC 423

RESULT 11
US-09-925-065A-788059
; Sequence 788059, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 788059
;   LENGTH: 630
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-788059

Alignment Scores:
Pred. No.:              0.479           Length:         630
Score:                  83.50           Matches:        36
Percent Similarity:     47.24%          Conservative:   24
Best Local Similarity:  28.35%          Mismatches:     28
Query Match:            15.07%          Indels:         39
DB:                     4               Gaps:           10

US-09-837-344-19 (1-107) x US-09-925-065A-788059 (1-630)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLys---------------------- 12
              |||             |||:::::::||||||:::|||
Db        177 CGAAGCCAAAGCAAGAAGAGAAGTTTAGAGAAAAAAGGGTGAAAAGAAATGAACAAAGCC 236

Qy         13 ---LysGluHisGlyAspIleLeuAlaGluAspLeuTyr--------------------- 24
                 |||:::::|||   :::      :::   :::|||
Db        237 TCCAAGAAATATGGGACTATGTGAAAAGACCAAATCTACATTTGATTAGTGTACCTGAAA 296

Qy         25 ------GlyArgLeuGluIleProAlaIleGluLeuProSer---GluAsnGluArgGly 41
                   ||||||:::|||                    ||||||   |||:::      |||
Db        297 GTGATGGGGAGAATGGAA-----------------CCAAGTTGRGAACACTCTTTAGGA 338

Qy         42 TyrTyr---------IleProHisGlnSerSerLeuProGlnAsp---AsnArgGlyAsn 57
              ||||||         :::|||      |||          ||||||   |||   ||||||
Db        339 TATTATCCAGGAGAACTTCCCCAACCTAGCAAGGCAGGCCAAGATTCAAATTCAGGAAAC 398

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
              :::   ::::::   ::::::   ::::::||||||   |||   :::      |||
Db        399 ACAGAGAACACTACAAAGCTA---CTACTCGAGAAGGGCAACGCCAAG------ACACAT 449

Qy         78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
              |||   :::           ||||||::          |||   |||         :::
Db        450 AATTGTCAG-----------ATTCACCAAAGTTGAAATGAAGGAAAA---------AAT 488

Qy         98 IleLysProGluGlnLysGlu 104
              :::|||      |||:::|||
Db        489 GTTAAGGGCAGCCAGAGAGAA 509

RESULT 12
US-09-925-065A-846536
; Sequence 846536, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 846536
;   LENGTH: 630
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-846536

Alignment Scores:
Pred. No.:              0.479           Length:         630
Score:                  83.50           Matches:        36
Percent Similarity:     47.24%          Conservative:   24
Best Local Similarity:  28.35%          Mismatches:     28
Query Match:            15.07%          Indels:         39
DB:                     4               Gaps:           10

US-09-837-344-19 (1-107) x US-09-925-065A-846536 (1-630)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLys---------------------- 12
              |||             |||:::::::||||||:::|||
Db        177 CGAAGCCAAAGCAAGAAGAGAAGTTTAGAGAAAAAAGGGTGAAAAGAAATGAACAAAGCC 236

Qy         13 ---LysGluHisGlyAspIleLeuAlaGluAspLeuTyr--------------------- 24
                 |||:::::|||   :::      :::   :::|||
Db        237 TCCAAGAAATATGGGACTATGTGAAAAGACCAAATCTACATTTGATTAGTGTACCTGAAA 296

Qy         25 ------GlyArgLeuGluIleProAlaIleGluLeuProSer---GluAsnGluArgGly 41
```

```
                 ||||||:::|||                    ||||||   |||:::      |||
Db        297 GTGATGGGGAGAATGGAA-----------------CCAAGTTGGGAACACTCTTTAGGA 338

Qy         42 TyrTyr---------IleProHisGlnSerSerLeuProGlnAsp---AsnArgGlyAsn 57
              ||||||         :::|||      |||         ||||||   |||   ||||||
Db        339 TATTATCCAGGAGAACTTCCCCAACCTAGCAAGGCAGGCCAAGATTCAAATTCAGGAAAC 398

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
              :::   ::::::   ::::::   ::::::||||||   |||   :::      |||
Db        399 ACAGAGAACACTACAAAGCTA---CTACTCGAGAAGGGCAACGCCAAG------ACACAT 449

Qy         78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
              |||   :::            ||||||:::         |||   |||         :::
Db        450 AATTGTCAG-----------ATTCACCAAAGTTGAAATGAAGGAAAA---------AAT 488

Qy         98 IleLysProGluGlnLysGlu 104
              :::|||      |||:::|||
Db        489 GTTAAGGGCAGCCAGAGAGAA 509

RESULT 13
US-09-922-261-193
; Sequence 193, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 193
;   LENGTH: 696
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-261-193

Alignment Scores:
Pred. No.:              1.01          Length:         696
Score:                  81.50         Matches:        31
Percent Similarity:     44.34%        Conservative:   16
Best Local Similarity:  29.25%        Mismatches:     54
Query Match:            14.71%        Indels:         5
DB:                     3             Gaps:           3

US-09-837-344-19 (1-107) x US-09-922-261-193 (1-696)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        232 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 291

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                    :::   |||             |||      :::   ||||||:::|||
Db        292 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 351

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||      |||
Db        352 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 411

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||             ||||||         :::         |||   |||
Db        412 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 471

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||      :::
Db        472 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 531

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        532 AAAGAAGATGAAAAAGAG 549

RESULT 14
US-09-922-261-191
; Sequence 191, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 191
;   LENGTH: 699
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-261-191

Alignment Scores:
Pred. No.:              1.01          Length:         699
Score:                  81.50         Matches:        31
Percent Similarity:     44.34%        Conservative:   16
Best Local Similarity:  29.25%        Mismatches:     54
Query Match:            14.71%        Indels:         5
DB:                     3             Gaps:           3

US-09-837-344-19 (1-107) x US-09-922-261-191 (1-699)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        235 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 294

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                    :::   |||             |||      :::   ||||||:::|||
Db        295 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 354

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||      |||
Db        355 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 414

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||             ||||||         :::         |||   |||
Db        415 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 474

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||                 ||||||      :::|||   ||||||      :::
Db        475 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 534

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        535 AAAGAAGATGAAAAAGAG 552

RESULT 15
US-09-922-261-189
; Sequence 189, Application US/09922261
```

```
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 189
;   LENGTH: 717
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-261-189

Alignment Scores:
Pred. No.:                1.05          Length:          717
Score:                    81.50         Matches:         31
Percent Similarity:       44.34%        Conservative:    16
Best Local Similarity:    29.25%        Mismatches:      54
Query Match:              14.71%        Indels:          5
DB:                       3             Gaps:            3

US-09-837-344-19 (1-107) x US-09-922-261-189 (1-717)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        253 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 312

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||       :::   ||||||:::|||
Db        313 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 372

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                    :::                   :::|||   :::||||||      |||
Db        373 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 432

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::         |||   |||
Db        433 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 492

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||             ||||||      :::|||   ||||||      :::
Db        493 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 552

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        553 AAAGAAGATGAAAAAGAG 570

RESULT 16
US-09-922-261-187
; Sequence 187, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187
;   LENGTH: 774
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-261-187

Alignment Scores:
Pred. No.:                1.16          Length:          774
Score:                    81.50         Matches:         31
Percent Similarity:       44.34%        Conservative:    16
Best Local Similarity:    29.25%        Mismatches:      54
Query Match:              14.71%        Indels:          5
DB:                       3             Gaps:            3

US-09-837-344-19 (1-107) x US-09-922-261-187 (1-774)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        310 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 369

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||             |||       :::   ||||||:::|||
Db        370 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 429

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                    :::                   :::|||   :::||||||      |||
Db        430 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 489

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
              ||||||            ||||||         :::         |||   |||
Db        490 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 549

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                 |||             ||||||      :::|||   ||||||      :::
Db        550 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 609

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        610 AAAGAAGATGAAAAAGAG 627

RESULT 17
US-09-922-261-185
; Sequence 185, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 185
;   LENGTH: 819
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-922-261-185

Alignment Scores:
Pred. No.:              1.25          Length:         819
Score:                  81.50         Matches:        31
Percent Similarity:     44.34%        Conservative:   16
Best Local Similarity:  29.25%        Mismatches:     54
Query Match:            14.71%        Indels:         5
DB:                     3             Gaps:           3

US-09-837-344-19 (1-107) x US-09-922-261-185 (1-819)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        355 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 414

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||           |||       :::   ||||||:::|||
Db        415 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 474

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||       |||
Db        475 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 534

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
                ||||||            ||||||         :::          |||   |||
Db        535 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 594

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                  |||             ||||||     :::|||   ||||||      :::
Db        595 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 654

Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::|||||
Db        655 AAAGAAGATGAAAAAGAG 672

RESULT 18
US-09-769-736-71
; Sequence 71, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
;  APPLICANT: Microbial Technics Limited
;  APPLICANT:  Le Page, Richard WF
;  APPLICANT:  Wells, Jeremy M
;  APPLICANT:  Hanniffy, Sean B
;  TITLE OF INVENTION: Proteins
;  FILE REFERENCE: PWC/P21089wo
;  CURRENT APPLICATION NUMBER: US/09/769,736
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: GB 9816335.5
;  PRIOR FILING DATE: 1998-07-27
;  PRIOR APPLICATION NUMBER: US 60/125163
;  PRIOR FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
;   LENGTH: 1455
;   TYPE: DNA
;   ORGANISM: Streptococcus agalactiae
US-09-769-736-71

Alignment Scores:
Pred. No.:              2.69          Length:         1455
Score:                  81.50         Matches:        22
Percent Similarity:     51.95%        Conservative:   18
Best Local Similarity:  28.57%        Mismatches:     20
Query Match:            14.71%        Indels:         17
DB:                     3             Gaps:           3

US-09-837-344-19 (1-107) x US-09-769-736-71 (1-1455)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAsp---------------------- 53
                ||||||:::   |||   :::   ||||||::::::
Db       1123 TACTACATGGAACATCCTAATGAATTACCAAAAGAAAAGGGTTGGGGACACGACAGTGAT 1182

Qy         54 ---AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle------GluLysThr 70
                  |||:::|||::::::::::|||:::|||      :::            |||:::
Db       1183 CATAACAAAGGCTCAAATAAAGACAATAAAGCCAAAAATTATGCTCCAGATGAAGAACCT 1242

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90
                              |||:::      |||      |||::::::||||||
Db       1243 GAAGATTCAGGGAAAGTAACTCACAACTATGGTTTTTATGATGTTAATAAAGGTTCAGAC 1302

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||                  :::||||||::::::||||||:::|||
Db       1303 GAAGAA-----------------GAACCAGAAAAACAAGAAGATGAATCA 1335

RESULT 19
US-09-922-261-184
; Sequence 184, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT:  Lo, Donald C.
;  APPLICANT:  Barney, Shawn
;  APPLICANT:  Thomas, Mary Beth
;  APPLICANT:  Portbury, Stuart D.
;  APPLICANT:  Puranam, Kasturi
;  APPLICANT:  Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION:  AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION:  CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/922,261
;  CURRENT FILING DATE:  2001-08-03
;  PRIOR APPLICATION NUMBER: US/09/461,697
;  PRIOR FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 184
;   LENGTH: 1669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-261-184

Alignment Scores:
Pred. No.:              3.23          Length:         1669
Score:                  81.50         Matches:        31
Percent Similarity:     44.34%        Conservative:   16
Best Local Similarity:  29.25%        Mismatches:     54
Query Match:            14.71%        Indels:         5
DB:                     3             Gaps:           3

US-09-837-344-19 (1-107) x US-09-922-261-184 (1-1669)

Qy          4 AspThrLysLysAsnLeuGlu---ArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||::::::   |||   :::   :::|||   |||:::         ||||||
Db        434 GATCAGAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAAGAAGAC 493

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
                  :::   |||           |||       :::   ||||||:::|||
Db        494 AAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAAGGAGAA 553

Qy         43 TyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                        :::                 :::|||   :::||||||       |||
Db        554 GATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGAGAAGAT 613

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal--- 79
                ||||||            ||||||         :::          |||   |||
Db        614 GGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGAGTTGGA 673

Qy         80 ---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                  |||             ||||||     :::|||   ||||||      :::
Db        674 AAAGAGAATGAAGATGGAAAAGAGAAGGGAGATAAAAAAGAGGGGAAAGATGTAAAAGTC 733
```

```
Qy         99 LysProGluGlnLysGlu 104
              |||   ::::::||||||
Db        734 AAAGAAGATGAAAAAGAG 751

RESULT 20
US-10-450-763-10274
; Sequence 10274, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 10274
;   LENGTH: 2460
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-450-763-10274

Alignment Scores:
Pred. No.:                5.41          Length:         2460
Score:                    81.50         Matches:        36
Percent Similarity:       43.08%        Conservative:   20
Best Local Similarity:    27.69%        Mismatches:     37
Query Match:              14.71%        Indels:         38
DB:                       9             Gaps:           6

US-09-837-344-19 (1-107) x US-10-450-763-10274 (1-2460)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLys------LysGluHisGlyAsp------ 17
              ||||||:::||||||   |||   |||   |||      :::   |||   :::
Db       1877 AAAGCAGAAACAAAGACAAATACAGAAACCAAGAGAGTAAGAGCTCACACAGAAAAGAGA 1936

Qy         18 ------------------------------------------IleLeuAlaGluAspLeu 23
                                                        |||:::|||:::
Db       1937 ATTCTGAGAGTGAGAAAAGAATGTACTCTAAAAGTCGTGATCATAATAGCTCAAATAACA 1996

Qy         24 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 43
                ||:::   :::   |||:::::||||||:::   :::
Db       1997 GCAGGGAAAAAAAGGCTGATAGAGATCAAAGTCCCTTCTCAAAAATA-AAA--------- 2046

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg-----------------GlyAsn 57
                 |||||||||      ||||||:::                          |||
Db       2047 ---------CAAAGCAGT------CAGGACGATGAATTAAAGTCCTCCATGTTGAAAAAT 2091

Qy         58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
                |||   |||     |||   :::||||||   |||::::::|||
Db       2092 AAGGAGGATGAGAAGATCAGATCCTCAGTGGAAAAAGAAAACCAAAAATCA--------- 2142

Qy         78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
                 :::|||:::   ::::::|||              :::|||   |||
Db       2143 ------AAAGGTCAAGAAAATGTCCATGTACATGAAAAAAATAAAAAATTTGATCATGAA 2196

Qy         98 IleLysProGluGlnLysGluAspLysSer 107
                |||         ||||||||||||
Db       2197 TCAAGCCCTGGAACAGATGAAGACAAAAGC 2226

RESULT 21
US-10-425-114-6320
; Sequence 6320, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 6320
;   LENGTH: 1508
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700560825_FLI
US-10-425-114-6320

Alignment Scores:
Pred. No.:                3.29          Length:         1508
Score:                    81.00         Matches:        26
Percent Similarity:       44.64%        Conservative:   24
Best Local Similarity:    23.21%        Mismatches:     32
Query Match:              14.62%        Indels:         30
DB:                       7             Gaps:           5

US-09-837-344-19 (1-107) x US-10-425-114-6320 (1-1508)

Qy          4 AspThrLysLysAsnLeuGlu--------------------------ArgLysLysGlu 14
              ||||||          :::|||                        ||||||:::
Db       1056 GATACAGCGATGGAAATAGAAGCACCAGTTGAAGTTTCAGGCAAGAAAGCCAAGAAGAAG 1115

Qy         15 HisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeu 34
              |||      ::::::   :::|||:::                        |||::::::
Db       1116 CATACAGCTGTCGTTACAGATGACATG--------------------GCTGTAGATAAG 1154

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn 54
              |||:::|||           |||      :::                  :::|||:::
Db       1155 CCCACAGAGACTACAAATGGTGATGCATTG--------------------GAGGATCAT 1193

Qy         55 ArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSer 74
              :::          :::      |||                  |||      |||:::
Db       1194 AAGTCAGAAAAGAAGAAGAGAAAG-----------------AAAGAAAAGAGGAAATTG 1235

Qy         75 IleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLys 94
                |||   |||   :::        |||          |||   :::|||   ::::::
Db       1236 GACACGGAGAACGATCAAGCCATGGATGATGGTGCTAATGGG---GTTGAAAGCGAACAA 1292

Qy         95 AspGlySerIleLysProGluGlnLysGluAspLys 106
              ||||||:::::::|||   ::::::|||:::||||||
Db       1293 GATGGAGCAGTTAAAAAGAAGAAAAAAAAAGGATAAA 1328

RESULT 22
US-10-424-599-91249/c
; Sequence 91249, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91249
;   LENGTH: 1689
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_53408C.1
```

```
US-10-424-599-91249

Alignment Scores:
Pred. No.:              3.82          Length:        1689
Score:                  81.00         Matches:         25
Percent Similarity:     48.57%        Conservative:    26
Best Local Similarity:  23.81%        Mismatches:      30
Query Match:            14.62%        Indels:          24
DB:                     7             Gaps:             4

US-09-837-344-19 (1-107) x US-10-424-599-91249 (1-1689)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGlu 21
              :::   :::   |||:::   :::|||   |||:::|||:::      :::   ||||||
Db       1673 GAAGGGGAGGTTAAGGAAACAATTGAAGGTAAACAAGAAAATGAAAAAGTGGAAGCTGAG 1614

Qy         22 AspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGly 41
                    :::   |||:::   |||:::|||:::
Db       1613 ------------AAACCAGAAGTAGATGCTATGGAAGTA---------------------- 1587

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSer 61
                        :::|||   :::||||||          :::   ::::::
Db       1586 --------------GAGAGTGGCATTCCGCAGAAGGAGGGGAGTGATAAAAAGGAGAAA 1542

Qy         62 LysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGly 81
              |||::::::::::::   |||:::                              ||||||
Db       1541 AAAGATGTGGCGATGGAGGAGGAGGATGAAGAT--------------------GAGGGT 1503

Qy         82 ArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGlu 101
              :::   :::|||   |||         |||||||||  :::   |||   |||   |||
Db       1502 AAGGATAATATTGATAAGTCAAAAGAAGAAGAGAAGGCAGAAGATAGT---AAGGGTGAG 1446

Qy        102 GlnLysGluAspLys 106
              ::::::      |||
Db       1445 AAAAGATCAAAAAAA 1431

RESULT 23
US-10-424-599-25309
; Sequence 25309, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25309
;   LENGTH: 1701
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_122856C.1
US-10-424-599-25309

Alignment Scores:
Pred. No.:              3.86          Length:        1701
Score:                  81.00         Matches:         26
Percent Similarity:     44.64%        Conservative:    24
Best Local Similarity:  23.21%        Mismatches:      32
Query Match:            14.62%        Indels:          30
DB:                     7             Gaps:             5

US-09-837-344-19 (1-107) x US-10-424-599-25309 (1-1701)

Qy          4 AspThrLysLysAsnLeuGlu-------------------------ArgLysLysGlu 14
              ||||||         :::|||                         ||||||:::
Db       1249 GATACAGCGATGGAAATAGAAGCACCAGTTGAAGTTTCAGGCAAGAAAGCCAAGAAGAAG 1308

Qy         15 HisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeu 34
              |||      ::::::   :::|||:::                        |||::::::
Db       1309 CATACAGCTGTCGTTACAGATGACATG--------------------GCTGTAGATAAG 1347

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn 54
              |||:::|||         |||   :::                       :::|||:::
Db       1348 CCCACAGAGACTACAAATGGTGATGCATTG--------------------GAGGATCAT 1386

Qy         55 ArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSer 74
              :::         :::     |||                   |||      |||:::
Db       1387 AAGTCAGAAAAGAAGAAGAGAAAG-----------------AAAGAAAAGAGGAAATTG 1428

Qy         75 IleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLys 94
                 |||   |||   :::         |||         |||   :::|||   ::::::
Db       1429 GACACGGAGAACGATCAAGCCATGGATGATGGTGCTAATGGG---GTTGAAAGCGAACAA 1485

Qy         95 AspGlySerIleLysProGluGlnLysGluAspLys 106
              ||||||::::::|||   ::::::|||:::||||||
Db       1486 GATGGAGCAGTTAAAAAGAAGAAAAAAAAAGGATAAA 1521

RESULT 24
US-10-357-930-49495
; Sequence 49495, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007BCN
;  CURRENT APPLICATION NUMBER: US/10/357,930
;  CURRENT FILING DATE:  2003-02-04
;  PRIOR APPLICATION NUMBER: 09/785,276
;  PRIOR FILING DATE: 2003-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49495
;   LENGTH: 540
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-357-930-49495

Alignment Scores:
Pred. No.:              0.977         Length:        540
Score:                  80.50         Matches:        37
Percent Similarity:     42.31%        Conservative:   18
Best Local Similarity:  28.46%        Mismatches:     38
Query Match:            14.53%        Indels:         38
DB:                     8             Gaps:            6

US-09-837-344-19 (1-107) x US-10-357-930-49495 (1-540)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLys------LysGluHisGlyAsp------ 17
              ||||||::||||||   |||   |||   |||      :::   |||   :::
Db        167 AAAGCAGAAACAAAGACAAATACAGAAACCAAGAGAGTAAGAGCTCACACAGAAAAGAAA 226

Qy         18 ----------------------------------------IleLeuAlaGluAspLeu 23
                                                       |||:::|||:::
Db        227 ATTCTGAGAGTGAGAAAAGAATGTACTCTAAAAGTCGTGATCATAATAGCTCAAATAACA 286
```

```
Qy          24 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 43
               |||:::      :::      |||::::::||||||:::      :::
Db         287 GCAGGGAAAAAAAGGCTGATAGAGATCAAAGTCCCTTCTCAAAAATA-AAA--------- 336

Qy          44 IleProHisGlnSerSerLeuProGlnAspAsnArg-----------------GlyAsn 57
                 |||||||||      |||||||||                          |||
Db         337 ---------CAAAGCAGT------CAGGACAATGAATTAAAGTCCTCCATGTTGAAAAAT 381

Qy          58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
                |||   |||      |||   :::||||||   |||:::::|||
Db         382 AAGGAGGATGAGAAGATCAGATCCTCAGTGGAAAAAGAAAACCAAAAATCA--------- 432

Qy          78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
                   :::|||:::   :::   |||            :::|||   |||
Db         433 ------AAAGGTCAAGAAAATGACCATGTACATGAAAAAAATAAAAAATTTGATCATGAA 486

Qy          98 IleLysProGluGlnLysGluAspLysSer 107
                  |||         ||||||||||||
Db         487 TCAAGCCCTGGAACAGATGAAGACAAAAGC 516

RESULT 25
US-09-925-065A-846535
; Sequence 846535, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 846535
;   LENGTH: 630
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-846535

Alignment Scores:
Pred. No.:              1.2          Length:        630
Score:                  80.50        Matches:       36
Percent Similarity:     46.46%       Conservative:  23
Best Local Similarity:  28.35%       Mismatches:    29
Query Match:            14.53%       Indels:        39
DB:                     4            Gaps:          10

US-09-837-344-19 (1-107) x US-09-925-065A-846535 (1-630)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLys----------------------- 12
               |||             |||::::::||||||:::|||
Db         177 CGAAGCCAAAGCAAGAAGAGAAGTTTAGAGAAAAAAGGGTGAAAAGAAATGAACAAAGCC 236

Qy          13 ---LysGluHisGlyAspIleLeuAlaGluAspLeuTyr--------------------- 24
                  |||::::::|||   :::      :::   :::|||
Db         237 TCCAAGAAATATGGGACTATGTGAAAAGACCAAATCTACATTTGATTAGTGTACCTGAAA 296

Qy          25 ------GlyArgLeuGluIleProAlaIleGluLeuProSer---GluAsnGluArgGly 41
                    ||||||:::|||                  ||||||   |||:::      |||
Db         297 GTGATGGGGAGAATGGAA-----------------CCAAGTTGGGAACACTCTTTAGGA 338

Qy          42 TyrTyr---------IleProHisGlnSerSerLeuProGlnAsp---AsnArgGlyAsn 57
                  ||||||          :::|||      |||          ||||||   |||   ||||||
Db         339 TATTATCCAGGAGAACTTCCCCAACCTAGCAAGGCAGGCCAAGATTCAAATTCAGGAAAC 398

Qy          58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
                ::: ::::::  ::::::  ::::::|||||  |||  :::      |||
Db         399 ACAGAGAACACTACAAAGCTA---CTACTCGAGAAGGGCAACGCCAAG------ACACAT 449

Qy          78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
                |||   :::         ||||||:::         |||   |||      :::
Db         450 AATTGTCAG-----------ATTCACCAAAGTTGAAATGAAGGAAAA---------AAT 488

Qy          98 IleLysProGluGlnLysGlu 104
                :::|||      |||   |||
Db         489 GTTAAGGGCAGCCAGASAGAA 509

RESULT 26
US-08-781-986A-521
; Sequence 521, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  521:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1353 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-521

Alignment Scores:
Pred. No.:              3.32         Length:        1353
Score:                  80.50        Matches:       30
Percent Similarity:     45.38%       Conservative:  24
Best Local Similarity:  25.21%       Mismatches:    39
Query Match:            14.53%       Indels:        26
DB:                     2            Gaps:          4

US-09-837-344-19 (1-107) x US-08-781-986A-521 (1-1353)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
               :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db         461 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 520

Qy          21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
               ::::::   :::      ::::::                   ||||||   :::|||
```

```
Db        521 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 580

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::           |||                    :::
Db        581 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 640

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                    ||| |||:::      :::|||:::   ||||||:::   |||::::::
Db        641 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 700

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                                        :::|||   |||
Db        701 CACAAAC----------------------------------------CTGCATCTCAAGAA 721

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 |||      ||||||         ||||||   ::::::||||||   :::
Db        722 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 776

RESULT 27
US-10-329-624-521
; Sequence 521, Application US/10329624
; Publication No. US20040043037A1
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/329,624
;             FILING DATE: 27-Dec-2002
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/956,171
;             FILING DATE: October 20, 1997
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1D1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 521:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1353 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 521:
US-10-329-624-521

Alignment Scores:
Pred. No.:              3.32          Length:        1353
Score:                  80.50         Matches:       30
Percent Similarity:     45.38%        Conservative:  24
Best Local Similarity:  25.21%        Mismatches:    39
Query Match:            14.53%        Indels:        26
DB:                     7             Gaps:          4

US-09-837-344-19 (1-107) x US-10-329-624-521 (1-1353)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        461 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 520

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::      ::::::                    ||||||   :::|||
Db        521 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 580

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::           |||                    :::
Db        581 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 640

Qy         50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
                    ||| |||:::      :::|||:::   ||||||:::   |||::::::
Db        641 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 700

Qy         69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
              ||||||                                        :::|||   |||
Db        701 CACAAAC----------------------------------------CTGCATCTCAAGAA 721

Qy         88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 |||      ||||||         ||||||   ::::::||||||   :::
Db        722 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 776

RESULT 28
US-10-381-596A-3
; Sequence 3, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
;  APPLICANT: Biostapro AB
;  TITLE OF INVENTION: von Willebrand factor-binding proteins from
;  TITLE OF INVENTION:  Staphylococci
;  FILE REFERENCE: 110059600
;  CURRENT APPLICATION NUMBER: US/10/381,596A
;  CURRENT FILING DATE:  2003-07-02
;  PRIOR APPLICATION NUMBER: SE 0003573-3
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1524
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-381-596A-3

Alignment Scores:
Pred. No.:              3.88          Length:        1524
Score:                  80.50         Matches:       30
Percent Similarity:     45.38%        Conservative:  24
Best Local Similarity:  25.21%        Mismatches:    39
Query Match:            14.53%        Indels:        26
DB:                     7             Gaps:          4

US-09-837-344-19 (1-107) x US-10-381-596A-3 (1-1524)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::||||||      |||:::   ||||||   |||||||||      :::   :::
Db        631 AAAAAAGCAGTTAACAAAAGGATGTTAGAAAATAAAAAAGAAGACTTAGAAACCATAATT 690

Qy         21 GluAspLeuTyrGlyArgLeuGlu-----------------IleProAlaIleGluLeu 34
              ::::::   :::      ::::::                    ||||||   :::|||
Db        691 GATGAATTTTTTAGTGATATAGATAAAACAAGACCTAATAATATTCCTGTTTTAGAAGAT 750

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGln--------------SerSer 49
                 :::   |||:::   :::           |||                    :::
Db        751 GAAAAACAAGAAGAGAAAAATCATAAAAATATGGCTCAATTAAAATCTGACACTGAAGCA 810
```

```
Qy          50 LeuProGlnAsp-AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLy 69
Db         811 GCAAAAAGTGATGAATCAAAAAGAAGCAAGAGAAGTAAAAGAAGTTTAAATACTCAAAAT 870

Qy          69 sThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis---LysGl 88
Db         871 CACAAAC-----------------------------------CTGCATCTCAAGAA 891

Qy          88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
Db         892 GTTTCTGAACAACAAAAAGCTGAATATGATAAAAGAGCAGAAGAAAGAAAAGCGA 946
```

RESULT 29
US-10-282-122A-16144
; Sequence 16144, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16144
; LENGTH: 2709
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-282-122A-16144

Alignment Scores:

Pred. No.:	8.35	Length:	2709
Score:	80.50	Matches:	28
Percent Similarity:	46.43%	Conservative:	24
Best Local Similarity:	25.00%	Mismatches:	43
Query Match:	14.53%	Indels:	17
DB:	7	Gaps:	5

US-09-837-344-19 (1-107) x US-10-282-122A-16144 (1-2709)

```
Qy           6 LysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGly 25
Db         814 GAGAAAAATAAACAGAGCATCAAGGAGCATAATGAATTAAGGGAAGAAACA--------- 864

Qy          26 ArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIlePro 45
Db         865 AAGAAAAAGCTTCCTAAGGTAAATATTGAGTTAAAGGAAGAAACAAAGAAACAAGTTCCT 924

Qy          46 HisGlnSer---SerLeuProGlnAsp--------------------AsnArgGlyAsn 57
Db         925 AATAAAAGTCTTGAATTAGATGAGGAAAAAGATAAACAAGTTTTCAAGGAACACAATGAA 984

Qy          58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
Db         985 TTAAGAAATGAGAAGAGTAAGCAGATTATCGAGGAACATAATGAA------------TCA 1032

Qy          78 AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
Db        1033 AAGGATGAAAAAAGCAAACAGGTTTCTAAGGGGAATATTGAATTAAAGAAAGAAAAAAGT 1092

Qy          98 ------IleLysProGluGlnLysGluAspLysSer 107
Db        1093 AAACAAATTATAAAGGAACACAATGAATCAAAGAAT 1128
```

RESULT 30
US-10-104-047-1957
; Sequence 1957, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1957
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1957

Alignment Scores:

Pred. No.:	5.17	Length:	1685
Score:	80.00	Matches:	29
Percent Similarity:	43.24%	Conservative:	19
Best Local Similarity:	26.13%	Mismatches:	53
Query Match:	14.44%	Indels:	10
DB:	6	Gaps:	3

US-09-837-344-19 (1-107) x US-10-104-047-1957 (1-1685)

```
Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
Db         707 AGAAGAAGATCAAAAGAAGATGAAGAAGATCAAAACGAAGAGAAAGGGGAAGCTGGAAAA 766

Qy          21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
Db         767 GAAGACAAAGATGAAAAAGGGGAAGAAGATGGAAAAGAGGATAAAAATGGAAATGAGAAA 826

Qy          41 GlyTyrTyrIleProHisGln------SerSerLeuProGlnAspAsnArgGlyAsnSer 58
Db         827 GGAGAAGATGCAAAAGAGAAAGAAGATGGAAAAAAAGGTGAAGACGGAAAAGGAAATGGA 886

Qy          59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr--- 77
Db         887 GAAGATGGAAAAGAGAAAGGAGAAGATGAAAAAGAGGAAGAAGACAGAAAAGAAACAGGG 946

Qy          78 -----------AsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLys 93
Db         947 GATGGAAAAGAGAATGAAGATGGAAAA---------GAGAAGGGAGATAAAAAAGAGGGG 997
```

```
Qy          94 LysAspGlySerIleLysProGluGlnLysGlu 104
               ||||||       :::|||   ::::::||||||
Db         998 AAAGATGTAAAAGTCAAAGAAGATGAAAAAGAG 1030
```

Search completed: December 20, 2005, 02:52:07
Job time : 701 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 02:52:18 ; Search time 214 Seconds
(without alignments)
245.090 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4161431 seqs, 245089505 residues

Total number of hits satisfying chosen parameters: 8322862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103710_335/app_query.fasta_1.263
-DB=Published_Applications_NA_New -QFMT=fastap -SUFFIX=rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_122_@runat_15122005_103710_335
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	79.5	14.4	8424	6	US-10-821-234-47	Sequence 47, Appl
c	2	78.5	14.2	2040	6	US-10-750-185-41854	Sequence 41854, A
	3	78	14.1	1344	6	US-10-793-626-2899	Sequence 2899, Ap
	4	78	14.1	1353	6	US-10-793-626-3225	Sequence 3225, Ap
c	5	78	14.1	3001	6	US-10-793-626-3813	Sequence 3813, Ap
	6	78	14.1	3974	6	US-10-793-626-4245	Sequence 4245, Ap
	7	76	13.7	193084	7	US-11-121-086-82	Sequence 82, Appl
	8	72	13.0	150468	7	US-11-112-908-56	Sequence 56, Appl
	9	72	13.0	193789	7	US-11-112-908-55	Sequence 55, Appl
	10	71.5	12.9	636	6	US-10-467-657-5485	Sequence 5485, Ap
	11	71.5	12.9	636	6	US-10-467-657-7403	Sequence 7403, Ap
c	12	71.5	12.9	648	6	US-10-793-626-595	Sequence 595, App
c	13	71.5	12.9	648	6	US-10-793-626-2257	Sequence 2257, Ap
c	14	71	12.8	984	6	US-10-467-657-6641	Sequence 6641, Ap
	15	71	12.8	984	6	US-10-467-657-7709	Sequence 7709, Ap
	16	70.5	12.7	451	7	US-11-108-172-1038	Sequence 1038, Ap
	17	70.5	12.7	120697	7	US-11-121-086-48	Sequence 48, Appl
	18	70.5	12.7	173602	7	US-11-121-086-25	Sequence 25, Appl
	19	69	12.5	162537	7	US-11-121-086-59	Sequence 59, Appl
	20	68.5	12.4	2081	7	US-11-196-475-71	Sequence 71, Appl
	21	68	12.3	2583	6	US-10-750-185-37927	Sequence 37927, A
	22	67.5	12.2	94618	7	US-11-117-187-191	Sequence 191, App
	23	67.5	12.2	403278	6	US-10-995-561-13421	Sequence 13421, A
	24	67	12.1	1107	6	US-10-467-657-1335	Sequence 1335, Ap
c	25	67	12.1	1116	6	US-10-467-657-1353	Sequence 1353, Ap
	26	67	12.1	1200	6	US-10-131-826A-273	Sequence 273, App
	27	67	12.1	1400	6	US-10-750-185-25422	Sequence 25422, A
	28	67	12.1	1605	6	US-10-750-185-60116	Sequence 60116, A
c	29	66	11.9	13083	6	US-10-995-561-13210	Sequence 13210, A
c	30	66	11.9	398287	6	US-10-995-561-13396	Sequence 13396, A
	31	65.5	11.8	2158	6	US-10-909-125-805	Sequence 805, App
	32	65.5	11.8	207835	7	US-11-121-086-39	Sequence 39, Appl
	33	65.5	11.8	207835	7	US-11-121-086-40	Sequence 40, Appl
	34	65.5	11.8	209822	6	US-10-995-561-13198	Sequence 13198, A
	35	65.5	11.8	317876	6	US-10-995-561-13227	Sequence 13227, A
	36	65	11.7	1611	7	US-11-186-284-114	Sequence 114, App
	37	65	11.7	2081	7	US-11-196-475-67	Sequence 67, Appl
	38	65	11.7	2126	7	US-11-196-475-75	Sequence 75, Appl
	39	64.5	11.6	1476	6	US-10-793-626-2807	Sequence 2807, Ap
c	40	64.5	11.6	1815	6	US-10-750-185-54973	Sequence 54973, A
	41	64.5	11.6	3857	6	US-10-793-626-3480	Sequence 3480, Ap
	42	64.5	11.6	19104	6	US-10-995-561-13425	Sequence 13425, A
c	43	64.5	11.6	91576	6	US-10-995-561-13461	Sequence 13461, A
c	44	64.5	11.6	165883	7	US-11-112-908-18	Sequence 18, Appl
	45	64	11.6	1974	6	US-10-467-657-4781	Sequence 4781, Ap

ALIGNMENTS

RESULT 1
US-10-821-234-47
; Sequence 47, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 47
; LENGTH: 8424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-47

Alignment Scores:

Pred. No.:	73.6	Length:	8424
Score:	79.50	Matches:	29
Percent Similarity:	42.20%	Conservative:	17
Best Local Similarity:	26.61%	Mismatches:	38
Query Match:	14.35%	Indels:	25
DB:	6	Gaps:	5

US-09-837-344-19 (1-107) x US-10-821-234-47 (1-8424)

```
Qy          7 LysAsnLeuGluArgLysLysGlu--------------HisGlyAspIleLeuAlaGlu 21
              |||:::::::||||||                      |||   :::||||||
Db       5267 AAGGACATGGAACGTCTTACAAGCCTTAGAGCTGGCATTGAAGGACCTTTACTTGCAAGT 5326

Qy         22 AspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGly 41
              |||:::   ||||||                                  :::|||
Db       5327 GATGTT---GGACGT-----------------------------------GACAGATCT 5347

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGln---------AspAsnArgGlyAsnSer 58
                    |||      ||||||:::|||:::         ::::::
Db       5348 GCTGCCAGCCCGGTTGTAAGTAGTATGCCAGAAAGAGCTTCAGAGTCTTCTTCAGAGGAA 5407

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              :::|||      ||||||   :::      |||          |||           |||
Db       5408 AAAGATGATTATGAAATTTTTGTAAAAGTTAAGGACACTCACGAAAAAAGCAAGAAAAAT 5467

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                        ::::::      ||||||      |||:::|||:::|||   |||:::
Db       5468 ------------AAGAACCGTGATAAGGGGGAGAAAGAGAAGAAAAAGAGACTCTTCATTA 5515

Qy         99 LysProGluGlnLysGluAspLysSer 107
              :::          |||      ||||||
Db       5516 AGATCTCGAAGTAAGCGTTCCAAATCT 5542
```

RESULT 2
US-10-750-185-41854/c
; Sequence 41854, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41854
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Bovine 19866881042089
US-10-750-185-41854

Alignment Scores:

Pred. No.:	15.3	Length:	2040
Score:	78.50	Matches:	21
Percent Similarity:	59.26%	Conservative:	11
Best Local Similarity:	38.89%	Mismatches:	11
Query Match:	14.17%	Indels:	11
DB:	6	Gaps:	3

US-09-837-344-19 (1-107) x US-10-750-185-41854 (1-2040)

```
Qy         17 AspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 36
              ::::::||||||   :::||||||||||         :::|||      |||      :::
Db       1259 GAGGTGCTTGCGAGAGAATTATATGGT---------CTGCCTTTTGCAGAGTGGCAAACT 1209

Qy         37 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 56
              |||:::|||||||||||                   ||||||||||||:::
Db       1208 GAAGATGAAAGAGGTTAC-----------------GACTTGCCCCAGGATCAT------ 1173

Qy         57 AsnSerArgAspSerLysGluIleSerIleIleGluLysThr 70
              :::||||||:::   |||   ::::::   |||      |||
Db       1172 AGTAGCAGGAATGGAAAAAGAATGGCAGAAATATTCAGCACA 1131
```

RESULT 3
US-10-793-626-2899
; Sequence 2899, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2899
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2899

Alignment Scores:

Pred. No.:	10.2	Length:	1344
Score:	78.00	Matches:	27
Percent Similarity:	43.12%	Conservative:	20
Best Local Similarity:	24.77%	Mismatches:	36
Query Match:	14.08%	Indels:	26
DB:	6	Gaps:	4

US-09-837-344-19 (1-107) x US-10-793-626-2899 (1-1344)

```
Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::|||||||||:::   ::::::
Db         82 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 141

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||        :::|||      |||
Db        142 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA---------------------- 174

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                           |||   |||        |||:::|||
Db        175 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 216

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              |||:::|||:::   :::      |||::::::     :::
Db        217 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 276

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              ||||||   |||                |||||||||   |||   :::
Db        277 CCAGAAGGAACTTCAGAT-----------------GAAGAAAAACCTGATGATGCACCA 318

Qy         99 LysProGluGlnLysGluAspLysSer 107
              ||||||:::   |||   :::::::::
Db        319 AAACCAGATGACAAACCTGAAGAAACA 345
```

RESULT 4
US-10-793-626-3225
; Sequence 3225, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3225
; LENGTH: 1353

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3225

Alignment Scores:
Pred. No.:                10.3          Length:          1353
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       6             Gaps:            4

US-09-837-344-19 (1-107) x US-10-793-626-3225 (1-1353)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db         82 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 141

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db        142 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 174

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                          |||   |||          |||:::|||
Db        175 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 216

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
                |||:::|||:::   :::      |||::::::      :::
Db        217 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 276

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              ||||||       |||                ||||||||   |||   :::
Db        277 CCAGAAGGAACTTCAGAT----------------GAAGAAAAACCTGATGATGCACCA 318

Qy         99 LysProGluGlnLysGluAspLysSer 107
              ||||||:::   |||   :::::::::
Db        319 AAACCAGATGACAAACCTGAAGAAACA 345

RESULT 5
US-10-793-626-3813/c
; Sequence 3813, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3813
;   LENGTH: 3001
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3813

Alignment Scores:
Pred. No.:                29.3          Length:          3001
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       6             Gaps:            4

US-09-837-344-19 (1-107) x US-10-793-626-3813 (1-3001)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db       2417 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 2358

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db       2357 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 2325

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                          |||   |||          |||:::|||
Db       2324 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 2283

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
                |||:::|||:::   :::      |||::::::      :::
Db       2282 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 2223

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              ||||||       |||                ||||||||   |||   :::
Db       2222 CCAGAAGGAACTTCAGAT----------------GAAGAAAAACCTGATGATGCACCA 2181

Qy         99 LysProGluGlnLysGluAspLysSer 107
              ||||||:::   |||   :::::::::
Db       2180 AAACCAGATGACAAACCTGAAGAAACA 2154

RESULT 6
US-10-793-626-4245
; Sequence 4245, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4245
;   LENGTH: 3974
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4245

Alignment Scores:
Pred. No.:                42.3          Length:          3974
Score:                    78.00         Matches:         27
Percent Similarity:       43.12%        Conservative:    20
Best Local Similarity:    24.77%        Mismatches:      36
Query Match:              14.08%        Indels:          26
DB:                       6             Gaps:            4

US-09-837-344-19 (1-107) x US-10-793-626-4245 (1-3974)

Qy          4 AspThrLysLysAsnLeuGlu--------------ArgLysLysGluHisGlyAspIle 18
              :::|||   :::||||||:::               :::||||||||||:::   ::::::
Db        499 GAAACTGCTCGTAATTTAAAAGCTGATATTGATGCTCAAAAGAAAGAGTACGAAGAACTC 558

Qy         19 LeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsn 38
                 |||   |||         :::|||      |||
Db        559 ------GAACAGCTTTCAAAAGAAATTGAAGCGTCAGCA-------------------- 591

Qy         39 GluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                          |||   |||          |||:::|||
Db        592 -----------------CCTAAACAAGATGAACCACCTAAAGATGAAGGTGCAGAAGTT 633

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
                |||:::|||:::   :::      |||::::::      :::
Db        634 GAAGATAACAAAGATGGTAATTCTGGAGAAGAATCAGAGAACAAACCTTCTGATGATGAA 693
```

```
Qy          79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
               ||||||      |||                 |||||||||   |||   :::
Db         694 CCAGAAGGAACTTCAGAT-----------------GAAGAAAAACCTGATGATGCACCA 735

Qy          99 LysProGluGlnLysGluAspLysSer 107
               ||||||:::   |||   :::::::::
Db         736 AAACCAGATGACAAACCTGAAGAAACA 762

RESULT 7
US-11-121-086-82
; Sequence 82, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 82
;   LENGTH: 193084
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-82

Alignment Scores:
Pred. No.:              1.22e+04        Length:         193084
Score:                  76.00           Matches:        35
Percent Similarity:     43.65%          Conservative:   20
Best Local Similarity:  27.78%          Mismatches:     25
Query Match:            13.72%          Indels:         46
DB:                     7               Gaps:           8

US-09-837-344-19 (1-107) x US-11-121-086-82 (1-193084)

Qy           6 LysLysAsnLeuGluArgLys--------------------------LysGluHisGly 16
               |||     ||||||:::|||                            |||:::::|||
Db      137195 AAGACAAAATTAGAGAAAAAGAATGAAAAGGAACAAACAAAGCCTCCAAGAAATATGGG 137254

Qy          17 AspIleLeuAlaGluAspLeuTyr--------------------------GlyArgLeu 27
                  :::     :::   :::|||                          ||||||:::
Db      137255 ACTATGTGAAAAGACCAAATTTACGTTTGATTGGAATACCTGAAAGTGATGGGAAGAATG 137314

Qy          28 GluIleProAlaIleGluLeuProSer--GluAsnGluArgGlyTyrTyrIleProHisG 47
               |||                  ||||||  ||||||     |||||||||
Db      137315 GAA-----------------CCAAGTTGGAAAACTCTTCAGGATATTAT---------C 137347

Qy          47 lnSerSerLeuProGln--------------------AspAsnArgGlyAsnSerArgA 60
                 |||||||||                         |||   ||||||:::   :
Db      137348 CAGGAGAACTTCCACAACCTAGAAAGACAGGCCAACATTCAAATTAAGGAAATACAGAGA 137407

Qy          60 spSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValG 80
               :::::      |||   ::::::||||||:::|||   ::::::
Db      137408 ACACCACAAACATA---CTCCTCGAGAAGAGCAACCCCAAGACA---------------- 137448

Qy          80 luGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysP 100
                     :::||||||:::|||      |||   |||         ::::::|||
Db      137449 --CATAATCATGAGATTCACCAAGGTCAAAATGAAGGAAAA---------AATGTTAAGG 137497

Qy         100 roGluGlnLysGlu 104
                 |||:::|||
Db      137498 GCAGCCAGAGAGAA 137511

RESULT 8
US-11-112-908-56
; Sequence 56, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
;  APPLICANT: Harris, Cole
;  APPLICANT:  Davis, Lisa M.
;  TITLE OF INVENTION: Breast Cancer Biomarkers
;  FILE REFERENCE: 04-164-US
;  CURRENT APPLICATION NUMBER: US/11/112,908
;  CURRENT FILING DATE:  2005-04-22
;  PRIOR APPLICATION NUMBER: US 60/564,758
;  PRIOR FILING DATE: 2004-04-23
;  PRIOR APPLICATION NUMBER: US 60/575,978
;  PRIOR FILING DATE: 2004-06-01
;  PRIOR APPLICATION NUMBER: US 60/631,702
;  PRIOR FILING DATE: 2004-11-30
;  PRIOR APPLICATION NUMBER: US 60/633,826
;  PRIOR FILING DATE: 2004-12-07
;  NUMBER OF SEQ ID NOS: 511
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 56
;   LENGTH: 150468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-112-908-56

Alignment Scores:
Pred. No.:              2.77e+04        Length:         150468
Score:                  72.00           Matches:        28
Percent Similarity:     40.57%          Conservative:   15
Best Local Similarity:  26.42%          Mismatches:     41
Query Match:            13.00%          Indels:         22
DB:                     7               Gaps:           4

US-09-837-344-19 (1-107) x US-11-112-908-56 (1-150468)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
               ||||||   :::   ||||||      ||||||||||||||||
Db      115628 AGAAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGGGAG---------AGA 115678

Qy          21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
               |||         ||||||   :::                          |||      :::
Db      115679 GAGAGGAAGGAAGGAAGGAGAAAG--------------------AAAGAAGAAAGGAAA 115717

Qy          41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
               |||                  :::   |||      :::      :::|||      |||
Db      115718 GGA---------AAGGAAGAAGGAAGTAAGGAAAAGAAAAGAAAAGGAAAAGAAAGAAAA 115768

Qy          61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
                  ||||||          :::|||      ::::::                 :::
Db      115769 GAGAAAGAAAGAAAGAAAGAAAAGAAAGAAAGAAAGAAAAGAAAGAGAGAAAGAAA 115828

Qy          81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
               ||||||                          |||:::|||:::|||      |||
Db      115829 GGGAGA--------------------------GAGAGGAAGGAAGGAAGGAGAAAGAAA 115861

Qy         101 GluGlnLysGluAspLys 106
               |||:::::::::   |||
Db      115862 GAAGAAAGGAAAGGAAAA 115879

RESULT 9
US-11-112-908-55
; Sequence 55, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
;  APPLICANT: Harris, Cole
;  APPLICANT:  Davis, Lisa M.
;  TITLE OF INVENTION: Breast Cancer Biomarkers
;  FILE REFERENCE: 04-164-US
;  CURRENT APPLICATION NUMBER: US/11/112,908
;  CURRENT FILING DATE:  2005-04-22
;  PRIOR APPLICATION NUMBER: US 60/564,758
;  PRIOR FILING DATE: 2004-04-23
;  PRIOR APPLICATION NUMBER: US 60/575,978
;  PRIOR FILING DATE: 2004-06-01
;  PRIOR APPLICATION NUMBER: US 60/631,702
;  PRIOR FILING DATE: 2004-11-30
```

```
;  PRIOR APPLICATION NUMBER: US 60/633,826
;  PRIOR FILING DATE: 2004-12-07
;  NUMBER OF SEQ ID NOS: 511
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 55
;   LENGTH: 193789
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-112-908-55
```

Alignment Scores:

Pred. No.:	3.86e+04	Length:	193789
Score:	72.00	Matches:	28
Percent Similarity:	40.57%	Conservative:	15
Best Local Similarity:	26.42%	Mismatches:	41
Query Match:	13.00%	Indels:	22
DB:	7	Gaps:	4

US-09-837-344-19 (1-107) x US-11-112-908-55 (1-193789)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||   :::   ||||||      ||||||||||||||
Db     165813 AGAAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGGGAG---------AGA 165863

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||         ||||||   :::                  |||      :::
Db     165864 GAGAGGAAGGAAGGAAGGAGAAAG--------------------AAAGAAGAAAGGAAA 165902

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              |||          :::   |||    :::     :::|||     |||
Db     165903 GGA---------AAGGAAGAAGGAAGTAAGGAAAAGAAAAGAAAAGGAAAAGAAAGAAAA 165953

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
                ||||||           :::|||     ::::::                 :::
Db     165954 GAGAAAGAAAGAAAGAAAGAAAAGAAAGAAAGAAAGAAAAGAAAGAAAGAGAGAAAGAAA 166013

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              ||||||                          |||:::|||:::|||      |||
Db     166014 GGGAGA-------------------------GAGAGGAAGGAAGGAAGGAGAAAGAAA 166046

Qy        101 GluGlnLysGluAspLys 106
              |||::::::::::   |||
Db     166047 GAAGAAAGGAAAGGAAAA 166064
```

```
RESULT 10
US-10-467-657-5485
; Sequence 5485, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5485
;   LENGTH: 636
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5485
```

Alignment Scores:

Pred. No.:	24.8	Length:	636
Score:	71.50	Matches:	35
Percent Similarity:	37.09%	Conservative:	21
Best Local Similarity:	23.18%	Mismatches:	38
Query Match:	12.91%	Indels:	57
DB:	6	Gaps:	8

US-09-837-344-19 (1-107) x US-10-467-657-5485 (1-636)

```
Qy          4 AspThrLysLysAsnLeuGluArgLysLysGlu-------------------------- 14
              |||:::||||||      |||:::|||::::::
Db        166 GACTCCAAAAAACTCAGCGAGAAAAAACGCGACGCGCTTGCCGAAATGATTAAAGAACAG 225

Qy         14 ------------------------------------------------------------ 14

Db        226 GCGGTCGCGTGGCACGTCGCCGCCTCCACGCCCGAAGAAATCGCAAGCCTCAACATCCTG 285

Qy         15 HisGlyAspIleLeuAla------GluAspLeuTyrGlyArgLeuGluIleProAlaIle 32
              |||      :::||||||           :::||||||
Db        286 CACGCCACCATGCTCGCCATGAAACGTGCCGTTTACGGT---------------------- 324

Qy         33 GluLeuProSerGluAsnGluArg-----------GlyTyrTyrIleProHisGlnSer 48
                 |||   :::      |||:::            |||      ||||||
Db        325 ---TTGGCTGCGCGTCCCGAAAAAATATTCATCGACGGCAACCGCATTCCCGAACATTTA 381

Qy         49 SerLeuProGlnAspAsn------ArgGlyAsnSerArg------AspSerLysGluIle 64
                 :::|||   :::         :::|||:::|||:::      :::
Db        382 GGCATCCCTGCCGAAGCAGTCGTCAAAGGCGACAGCAAAATCATCGAAATCTCCGCCGCA 441

Qy         65 SerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg--- 83
              |||::::::   ||||||   |||::::::           :::      ||||||
Db        442 TCCGTTTTGGCAAAGACCGCACGCGATGCGGAAATGTACGCACTGGCGCAACGCCGTCCC 501

Qy         84 ------------AspIleHisLysGly---------------HisLeuGluGluLysLys 94
                        |||   |||||||||               |||||||||      |||
Db        502 CAATACGGTTTCGACAAACACAAAGGTTACGGCACAAAGCAGCATCTGGAAGCCCTCAAA 561

Qy         95 AspGlySerIleLysProGluGlnLysGluAsp 105
                    :::   ||||||   :::   |||
Db        562 CAATACGGCGTGCTGCCCGAACACCGCCGCGAC 594
```

```
RESULT 11
US-10-467-657-7403
; Sequence 7403, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7403
;   LENGTH: 636
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7403
```

Alignment Scores:

Pred. No.:	24.8	Length:	636
Score:	71.50	Matches:	35
Percent Similarity:	37.09%	Conservative:	21
Best Local Similarity:	23.18%	Mismatches:	38
Query Match:	12.91%	Indels:	57
DB:	6	Gaps:	8

US-09-837-344-19 (1-107) x US-10-467-657-7403 (1-636)

```
Qy          4 AspThrLysLysAsnLeuGluArgLysLysGlu-------------------------- 14
              |||:::||||||      |||:::|||::::::
```

```
Db        166 GACTCCAAAAAACTCAGCGAGAAAAAACGCGACGCGCTTGCCGAAATGATTAAAGAACAG 225

Qy         14 ------------------------------------------------------------ 14

Db        226 GCGGTCGCGTGGCACGTCGCCGCCTCCACGCCCGAAGAAATCGCAAGCCTCAACATCCTG 285

Qy         15 HisGlyAspIleLeuAla------GluAspLeuTyrGlyArgLeuGluIleProAlaIle 32

Db        286 CACGCCACCATGCTCGCCATGAAACGTGCCGTTTACGGT---------------------- 324

Qy         33 GluLeuProSerGluAsnGluArg------------GlyTyrTyrIleProHisGlnSer 48

Db        325 ---TTGGCTGCGCGTCCCGAAAAAATATTCATCGACGGCAACCGCATTCCCGAACATTTA 381

Qy         49 SerLeuProGlnAspAsn------ArgGlyAsnSerArg------AspSerLysGluIle 64

Db        382 GGCATCCCTGCCGAAGCAGTCGTCAAAGGCGACAGCAAAATCATCGAAATCTCCGCCGCA 441

Qy         65 SerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg--- 83

Db        442 TCCGTTTTGGCAAAGACCGCACGCGATGCGGAAATGTACGCACTGGCGCAACGCCGTCCC 501

Qy         84 ------------AspIleHisLysGly---------------HisLeuGluGluLysLys 94

Db        502 CAATACGGTTTCGACAAACACAAAGGTTACGGCACAAAGCAGCATCTGGAAGCCCTCAAA 561

Qy         95 AspGlySerIleLysProGluGlnLysGluAsp 105

Db        562 CAATACGGCGTGCTGCCCGAACACCGCCGCGAC 594

RESULT 12
US-10-793-626-595/c
; Sequence 595, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 595
;   LENGTH: 648
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-595

Alignment Scores:
Pred. No.:              25.4          Length:         648
Score:                  71.50         Matches:        24
Percent Similarity:     42.27%        Conservative:   17
Best Local Similarity:  24.74%        Mismatches:     35
Query Match:            12.91%        Indels:         21
DB:                     6             Gaps:           3

US-09-837-344-19 (1-107) x US-10-793-626-595 (1-648)

Qy         11 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 30

Db        647 CAAAAGAAAGAGTACGAAGAACTC------GAACAGCTTTCAAAAGAAATTGAAGCGTCA 594

Qy         31 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 50

Db        593 GCA-----------------------------------CCTAAACAAGATGAACCA 573

Qy         51 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 70

Db        572 CCTAAAGATGAAGGTGCAGAAGTTGAAGATAACAAAGATGGTAATTCTGGAGAAGAATCA 513

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90

Db        512 GAGAACAAACCTTCTGATGATGAACCAGAAGGAACTTCAGAT---------------- 471

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107

Db        470 GAAGAAAAACCTGATGATGCACCAAAACCAGATGACAAACCTGAAGAAACA 420

RESULT 13
US-10-793-626-2257/c
; Sequence 2257, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2257
;   LENGTH: 648
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2257

Alignment Scores:
Pred. No.:              25.4          Length:         648
Score:                  71.50         Matches:        24
Percent Similarity:     42.27%        Conservative:   17
Best Local Similarity:  24.74%        Mismatches:     35
Query Match:            12.91%        Indels:         21
DB:                     6             Gaps:           3

US-09-837-344-19 (1-107) x US-10-793-626-2257 (1-648)

Qy         11 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 30

Db        647 CAAAAGAAAGAGTACGAAGAACTC------GAACAGCTTTCAAAAGAAATTGAAGCGTCA 594

Qy         31 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 50

Db        593 GCA-----------------------------------CCTAAACAAGATGAACCA 573

Qy         51 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 70

Db        572 CCTAAAGATGAAGGTGCAGAAGTTGAAGATAACAAAGATGGTAATTCTGGAGAAGAATCA 513

Qy         71 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 90

Db        512 GAGAACAAACCTTCTGATGATGAACCAGAAGGAACTTCAGAT---------------- 471

Qy         91 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107

Db        470 GAAGAAAAACCTGATGATGCACCAAAACCAGATGACAAACCTGAAGAAACA 420

RESULT 14
US-10-467-657-6641/c
; Sequence 6641, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
```

; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6641
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-6641

Alignment Scores:

Pred. No.:	50.8	Length:	984
Score:	71.00	Matches:	25
Percent Similarity:	49.38%	Conservative:	15
Best Local Similarity:	30.86%	Mismatches:	35
Query Match:	12.82%	Indels:	6
DB:	6	Gaps:	3

US-09-837-344-19 (1-107) x US-10-467-657-6641 (1-984)

```
Qy         25 GlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr--- 43
              |||::::::|||         :::|||:::     |||      |||||||||
Db        576 GGTCAAATTGAA---------ATGGAAATTAAAGTAGAGCAAGGTCGTGGTTATCAATCT 526

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg------GlyAsnSerArgAspSer 61
              :::     :::     :::     :::||||||     |||     :::
Db        525 GTTTCAGGTCGTCAGGTAGTTCGTGATGAGAACCGTCAGATTGGTGCAATCCAGTTGGAT 466

Qy         62 LysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGly 81
                  |||   |||   :::   :::   |||                |||||
Db        465 GCGAGCTTTTCGCCCATCAGCCGTGTTAGCTTTGAGGTTGAACCTGCACGTGTAGAGCAG 406

Qy         82 ArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGlu 101
              |||   |||:::   |||      |||:::   :::   ||||||||||||   ||||||
Db        405 CGGACGGATCTTGATAAGTTGGTTTTGGATATCGAAACCGACGGTTCTATTGATCCTGAG 346

Qy        102 Gln 102
              :::
Db        345 GAA 343
```

RESULT 15
US-10-467-657-7709
; Sequence 7709, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7709
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7709

Alignment Scores:

Pred. No.:	50.8	Length:	984
Score:	71.00	Matches:	25
Percent Similarity:	49.38%	Conservative:	15
Best Local Similarity:	30.86%	Mismatches:	35
Query Match:	12.82%	Indels:	6
DB:	6	Gaps:	3

US-09-837-344-19 (1-107) x US-10-467-657-7709 (1-984)

```
Qy         25 GlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr--- 43
              |||::::::|||         :::|||:::     |||      |||||||||
Db        409 GGTCAAATTGAA---------ATGGAAATTAAAGTAGAGCAAGGTCGTGGTTATCAATCT 459

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg------GlyAsnSerArgAspSer 61
              :::     :::     :::     :::||||||     |||     :::
Db        460 GTTTCAGGTCGTCAGGTAGTTCGTGATGAGAACCGTCAGATTGGTGCAATCCAGTTGGAT 519

Qy         62 LysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGly 81
                  |||   |||   :::   :::   |||                |||||
Db        520 GCGAGCTTTTCGCCCATCAGCCGTGTTAGCTTTGAGGTTGAACCTGCACGTGTAGAGCAG 579

Qy         82 ArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGlu 101
              |||   |||:::   |||      |||:::   :::   ||||||||||||   ||||||
Db        580 CGGACGGATCTTGATAAGTTGGTTTTGGATATCGAAACCGACGGTTCTATTGATCCTGAG 639

Qy        102 Gln 102
              :::
Db        640 GAA 642
```

RESULT 16
US-11-108-172-1038
; Sequence 1038, Application US/11108172
; Publication No. US20050260177A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/833,263
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 09/649,811
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/609,448
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/575,251
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,444
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 09/504,629
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US 09/480,321
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: US 09/476,296
; PRIOR FILING DATE: 1999-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1130

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1038
;   LENGTH: 451
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-108-172-1038

Alignment Scores:
Pred. No.:              21.1          Length:        451
Score:                  70.50         Matches:       24
Percent Similarity:     43.48%        Conservative:  16
Best Local Similarity:  26.09%        Mismatches:    33
Query Match:            12.73%        Indels:        19
DB:                     7             Gaps:          4

US-09-837-344-19 (1-107) x US-11-108-172-1038 (1-451)

Qy         28 GluIleProAlaIleGluLeuProSerGluAsnGluArg---------------------- 40
              |||:::|||   |||     |||::::::     |||
Db         14 GAGCTGCCGACCATCAGGACGCCTGCAGACATTTACAGAGCCTTTGTTGATGTTGTGAAT 73

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer------ 58
              |||   |||:::|||   :::|||   |||         :::|||   ||||||
Db         74 GGAGAATATGTCCCTCGCAAATCCATCCTGAAGTCTCGAAGTAGAGAGAATAGTGTGTGT 133

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
                |||:::   |||   |||       |||      :::|||      :::      :::
Db        134 AGCGACACTAGTGAAAGCAGTGCTGCTGAATTTGATGATAGGCGGGGAGTTTTGAGGAGT 193

Qy         79 ValGluGlyArgArg------------AspIleHisLysGlyHisLeuGluGluLysLys 94
              :::                      |||      :::      |||||||||
Db        194 ATCAGCTGCGAAGAAGCCACTTGCAGTGACACCAGTGAGAGCATTTTGGAAGAG------ 247

Qy         95 AspGlySerIleLysProGluGlnLysGluAspLys 106
                      :::|||::::::   :::   |||
Db        248 ------------GAACCACAAGAAAATCAAAAGAAA 271

RESULT 17
US-11-121-086-48
; Sequence 48, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 48
;   LENGTH: 120697
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-48

Alignment Scores:
Pred. No.:              3.19e+04      Length:        120697
Score:                  70.50         Matches:       33
Percent Similarity:     43.52%        Conservative:  14
Best Local Similarity:  30.56%        Mismatches:    50
Query Match:            12.73%        Indels:        11
DB:                     7             Gaps:          5

US-09-837-344-19 (1-107) x US-11-121-086-48 (1-120697)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||:::   :::   ||||||     ||||||:::||||||        ||||||:::
Db     101644 AGAAGGAAGGAGGGGAAGAAGGAA---GAAAGAAGGAAAGAAAGATTCATTATTCTGTCC 101700

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAla---IleGluLeuProSerGluAsnGlu 39
                  :::         |||   :::||||||      :::   |||      |||   :::
Db     101701 TCTGAACCTGAGTGAAGA---AAAATACCCTGTCCTTTGTACCTGCGTGAAGAGAGAAAA 101757

Qy         40 ArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArg 59
                                        :::            :::          |||
Db     101758 GAAAAGAAAAGAAAAGAAAGAAAGAAAGAAAAGAAAAGACAGGAAGGAAGAAAGGAA 101817

Qy         60 AspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal 79
                 ||||||                |||   :::|||:::
Db     101818 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAAAAAGACAGACAG----------------- 101859

Qy         80 GluGlyArg------ArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
              |||||||||      |||        ||||||      |||   ||||||:::|||
Db     101860 GAAGGAAGAAAGGAAAGAAAGAAAGAAAAGGGAAGGAAGGAAGGAAAGAAGGAAGGAAGG 101919

Qy         98 IleLysProGluGlnLysGluAsp 105
              |||      :::||||||||||
Db     101920 AAGAAAGAAGGAAAGAAAGAAGAT 101943

RESULT 18
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
;   LENGTH: 173602
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-25

Alignment Scores:
Pred. No.:              5.14e+04      Length:        173602
Score:                  70.50         Matches:       29
Percent Similarity:     40.38%        Conservative:  13
Best Local Similarity:  27.88%        Mismatches:    47
Query Match:            12.73%        Indels:        15
DB:                     7             Gaps:          3

US-09-837-344-19 (1-107) x US-11-121-086-25 (1-173602)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||   :::   ||||||      ||||||||||||||      :::
Db      56054 AGAAAGAAAGAAAGAAAGAAA------GAAAGAAAGAAAGAAAAGAAAGAAAGAAAGAAA 56107

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||          ||||||   |||                                 |||
Db      56108 GAAAGGAAGGAAGGAAGGAAGGAA--------------------GGAGGGAAGGAAGGA 56146

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              |||                  :::            ::::::   :::          |||
Db      56147 GGGAAGGAAGGAGGGAAGGAAGGAGAGAGAGAGAAAGAAAGAAAGAAAAGAAAGAGAAAG 56206

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              ||||||                          ||||||                   |||
Db      56207 AAGAAAAGAG-----------------GAAGAGAGAAAGAAAGAAAGAAAGAAAGGAA 56248

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100
              |||||||||:::      :::         :::|||:::|||:::          |||
Db      56249 GGAAGGAGGGAAGGAAGGAGAGAGAGAAAGAAAGAAAGAAAAGAAAGAAAGAGAAAGAAG 56308

Qy        101 GluGlnLysGlu 104
              ::::::||||||
Db      56309 AAAGAGAAAGAA 56320
```

```
RESULT 19
US-11-121-086-59
; Sequence 59, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 59
;   LENGTH: 162537
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-59

Alignment Scores:
Pred. No.:              7.26e+04        Length:         162537
Score:                  69.00           Matches:        16
Percent Similarity:     52.17%          Conservative:   8
Best Local Similarity:  34.78%          Mismatches:     14
Query Match:            12.45%          Indels:         8
DB:                     7               Gaps:           1

US-09-837-344-19 (1-107) x US-11-121-086-59 (1-162537)

Qy          11 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 30
               ||||||:::   ||||||   :::||||||:::|||      |||
Db      116346 AGGAAACGGCGGCATGGGGTGCTGCTGGCTCAGGATGGGGCAGGC--------------- 116390

Qy          31 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 50
                     |||       |||:::      |||   :::   ||||||      ::::::
Db      116391 ---------CTGGAACCTGAGCACGCTGTCGGGCCCTGGACCCCTCACCATCCTGCAGTG 116441

Qy          51 ProGlnAspAsnArgGly 56
               |||   :::   ||||||
Db      116442 CCAGGAGAGGGCAGGGGC 116459

RESULT 20
US-11-196-475-71
; Sequence 71, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 71
;   LENGTH: 2081
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2079)
US-11-196-475-71

Alignment Scores:
Pred. No.:              278             Length:         2081
Score:                  68.50           Matches:        27
Percent Similarity:     40.98%          Conservative:   23
Best Local Similarity:  22.13%          Mismatches:     43
Query Match:            12.36%          Indels:         29
DB:                     7               Gaps:           4

US-09-837-344-19 (1-107) x US-11-196-475-71 (1-2081)

Qy           2 LysAlaAspThrLysLysAsnLeuGluArgLys--------------LysGluHisGly 16
               ||||||   :::|||:::   ::::::   |||              ||||||
Db        1231 AAAGCTTCTAGTAAAGAAAAAATTAAAGGCAAAGAAAAAGAAATAGTCAAAGAGAAATCA 1290

Qy          17 AspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 36
                  :::         ||||||
Db        1291 AAGGTAAGTTTAGGTGATTTG----------------------------------------- 1311

Qy          37 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 56
               :::|||:::                |||            |||   :::|||   :::
Db        1312 GATAATGACGAAACCCTTATGACGCCGGAAGATCAAAAATTATCTGAGGATAAAAAATTA 1371

Qy          57 AsnSerArgAspSer---LysGluIleSerIleIleGluLysThrAsnArgGluSerIle 75
               :::|||:::   :::   |||   :::|||   ||||||:::   |||       |||
Db        1372 GATAGTAAAAAAAATTTAAAACCTGTTTCTGAGATTGAGAGAGTAAATGAAATTTCAAAG 1431

Qy          76 ThrThrAsnValGluGlyArgArg----------------------------AspIle 85
               :::   |||   |||       :::                           ||||||
Db        1432 TCTAACAACAATGAGGTTAGCAAATCATCACCATTAGATAAGCCTTCTTATAGTGATATC 1491

Qy          86 HisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAsp 105
                     :::   :::   ||||||   :::::::::       :::   :::
Db        1492 GATTCAAAAGAGGTTGTAGATAATAAAGATGTTAATTTGCAAGAAACCAAGCCTCAAGCT 1551

Qy         106 LysSer 107
               ||||||
Db        1552 AAAAGT 1557

RESULT 21
US-10-750-185-37927
; Sequence 37927, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 37927
;   LENGTH: 2583
;   TYPE: DNA
;   ORGANISM: Bovine  19866880881676
US-10-750-185-37927

Alignment Scores:
Pred. No.:              426             Length:         2583
Score:                  68.00           Matches:        17
```

```
Percent Similarity:     55.10%          Conservative:  10
Best Local Similarity:  34.69%          Mismatches:    18
Query Match:            12.27%          Indels:        4
DB:                     6               Gaps:          1

US-09-837-344-19 (1-107) x US-10-750-185-37927 (1-2583)

Qy         16 GlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu-----------GluIleProAla 31
              |||   :::|||          |||:::|||   |||                   |||:::
Db        275 GGCCACATGCTACTGATTATTTTGTTTGGCCACTTGGTTGACATAAGCAGTTATCCTTCC 334

Qy         32 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 51
                        |||::::::|||||||||      :::||||||:::||||||
Db        335 TGTTTACCAGACTCTGACAGTGAAAGAGGATCAAGTCTTCCCCATGAATCCTCCCAAATT 394

Qy         52 GlnAspAsnArgGlyAsnSerArgAsp 60
              :::   :::||||||:::      |||
Db        395 CGACAAGATCGAGGACATGGCCATGAT 421

RESULT 22
US-11-117-187-191
; Sequence 191, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
;  APPLICANT: PREUSS, DAPHNE
;  APPLICANT:  COPENHAVER, GREGORY
;  TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
;  FILE REFERENCE: ARCD:309US
;  CURRENT APPLICATION NUMBER: US/11/117,187
;  CURRENT FILING DATE:  2005-04-28
;  PRIOR APPLICATION NUMBER: US/09/531,120
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/125,219
;  PRIOR FILING DATE: 1999-03-18
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
;   LENGTH: 94618
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-11-117-187-191

Alignment Scores:
Pred. No.:              5.49e+04        Length:        94618
Score:                  67.50           Matches:       22
Percent Similarity:     44.16%          Conservative:  12
Best Local Similarity:  28.57%          Mismatches:    29
Query Match:            12.18%          Indels:        14
DB:                     7               Gaps:          2

US-09-837-344-19 (1-107) x US-11-117-187-191 (1-94618)

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSer 61
              ||||||   |||   :::          |||      |||         ||||||:::
Db      75720 TACTACAGCCCACCACGAAAAAAACGTCCAAGTAGAAATAGTGAAGGAAGTAGAGAAGGT 75779

Qy         62 Lys-GluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn-------- 78
              ::: ::::::|||:::           :::   |||   :::|||   |||
Db      75780 AGGAGATGTTTCCGTT---------AAGGATACTGAGAAGGTAACCCAAAACAAAACTCT 75830

Qy         79 ----------------------ValGluGlyArgArgAspIleHisLysGlyHisLeuGl 91
                                   |||||||||      |||      :::|||   |||||
Db      75831 ATCATCGATGCAAAAAGTTGATGTTGAAGGGATGGGAGACAAGGAGGAAGGTACTCTTGA 75890

Qy         91 uGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              |   ::::::||||||                        |||      :::|||
Db      75891 AATTGAGCAGGATGGAGGATGGATTACTCCTTCTAAGACTGGTCGTAGT 75939

RESULT 23
US-10-995-561-13421
; Sequence 13421, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13421
;   LENGTH: 403278
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(403278)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13421

Alignment Scores:
Pred. No.:              3.59e+05        Length:        403278
Score:                  67.50           Matches:       35
Percent Similarity:     46.27%          Conservative:  27
Best Local Similarity:  26.12%          Mismatches:    26
Query Match:            12.18%          Indels:        46
DB:                     6               Gaps:          11

US-09-837-344-19 (1-107) x US-10-995-561-13421 (1-403278)

Qy          5 ThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyr 24
              :::||||||||   :::      |||:::::::|||   :::      :::   :::|||
Db     321783 AGTAAAAAGAAC---AAAGCCTCCAAGAAATATGGGACTATGTGAAAAGACCAAATCTAC 321839

Qy         25 --------------------------GlyArgLeuGluIleProAlaIleGluLeuPro 35
                                         ||||||:::|||                 |||
Db     321840 ATTTGATTGGTGTACCTGAAAGTCACGGGGAGAATGGAA-----------------CCA 321881

Qy         36 Ser----GluAsnGluArgGlyTyrTyr---------IleProHisGlnSerSerLeuPr 51
              |||    ::::::      |||||||||         :::|||      |||
Db     321882 AGTTAGAAAACACTCTTCAGGATATTATCCAGGAGAACTTCCCCAACATAGCAAGGCAGG 321941

Qy         51 oGln---AspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysTh 70
               |||      |||   ||||||:::   ::::::   :::|||   :::::::::|||::
Db     321942 CCAACATTCAAATTCAGGAAATACAGAGAACACCACAAAGATA---CTCCTCAAGAAGAG 321998

Qy         70 rAsnArgGluSerIleThrThrAsnVal--------------GluGlyArgArgAspIl 85
              :|||   ::::::      :::   :::                |||||:::
Db     321999 CAACMCCAAGACACATAATTCTCAGATTCACCAAGGTTGAAATGAAGGAAAAAAT----- 322053

Qy         85 eHisLysGlyHisLeuGluGluLysLysAsp---------------------------- 95
              ||||||   |||   |||:::   |||
Db     322054 -GTTAAGGGTAGCCTGAGAGAAAGGTCAGATTACCCACTAACRGAAGCCCATCAGACTAA 322112

Qy         96 ----GlySerIle--------------LysProGluGln 102
                  |||||||:::              |||||||||:::
Db     322113 CAGCGGCTCTCTCGGGCAGAAATCATACAAGCCAGAAGAG 322152

RESULT 24
US-10-467-657-1335
; Sequence 1335, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
```

```
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1335
;   LENGTH: 1107
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1335

Alignment Scores:
Pred. No.:                187             Length:         1107
Score:                    67.00           Matches:        20
Percent Similarity:       46.58%          Conservative:   14
Best Local Similarity:    27.40%          Mismatches:     29
Query Match:              12.09%          Indels:         10
DB:                       6               Gaps:           1

US-09-837-344-19 (1-107) x US-10-467-657-1335 (1-1107)

Qy          30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
               |||          :::   ||||||   :::|||||      :::|||         :::
Db         487 CCGGGTCATACCGTACGCAGTGAATACAAACGCGGCGGCGGCGTGCCTTCTCTGATTGCC 546

Qy          50 LeuProGlnAspAsnArgGlyAsnSerArgAsp---------------------------- 60
               :::   |||||||||   |||   ::::::|||
Db         547 GTTTACCAAGACAATTCCGGCAAGGCCAAAGACATCGCCCTGTCTTATGCGGCTGCCAAC 606

Qy          61 ---SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal 79
                  :::      :::||||||   |||   ||||||          |||::::::
Db         607 GGCGGCACCAAAGGCGGTGTGATTGAAACCACTTTCCGCGAAGAAACCGAAACCGATCTG 666

Qy          80 GluGlyArgArgAspIleHisLysGlyHisLeuGluGlu 92
               |||   :::   :::       |||   :::   |||
Db         667 TTCGGCGAACAAGCCGTATTGTGCGGTGGCGTGGCCGAG 705

RESULT 25
US-10-467-657-1353/c
; Sequence 1353, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1353
;   LENGTH: 1116
;   TYPE: DNA
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1353

Alignment Scores:
Pred. No.:                189             Length:         1116
Score:                    67.00           Matches:        20
Percent Similarity:       46.58%          Conservative:   14
Best Local Similarity:    27.40%          Mismatches:     29
Query Match:              12.09%          Indels:         10
DB:                       6               Gaps:           1

US-09-837-344-19 (1-107) x US-10-467-657-1353 (1-1116)

Qy          30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSer 49
               |||          :::   ||||||   :::|||||||      :::|||         :::
Db         543 CCGGGTCATACCGTACGCAGTGAATACAAACGCGGCGGCGGCGTGCCTTCTCTGATTGCC 484

Qy          50 LeuProGlnAspAsnArgGlyAsnSerArgAsp---------------------------- 60
               :::   |||||||||   |||   ::::::|||
Db         483 GTTTACCAAGACAATTCCGGCAAGGCCAAAGACATCGCCCTGTCTTATGCGGCTGCCAAC 424

Qy          61 ---SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnVal 79
                  :::      :::||||||   |||   ||||||          |||::::::
Db         423 GGCGGCACCAAAGGCGGTGTGATTGAAACCACTTTCCGCGAAGAAACCGAAACCGATCTG 364

Qy          80 GluGlyArgArgAspIleHisLysGlyHisLeuGluGlu 92
               |||   :::   :::       |||   :::   |||
Db         363 TTCGGCGAACAAGCCGTATTGTGCGGTGGCGTGGCCGAG 325

RESULT 26
US-10-131-826A-273
; Sequence 273, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C128
;  CURRENT APPLICATION NUMBER: US/10/131,826A
;  CURRENT FILING DATE:  2002-04-24
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 273
;   LENGTH: 1200
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-131-826A-273

Alignment Scores:
Pred. No.:                208             Length:         1200
Score:                    67.00           Matches:        21
Percent Similarity:       40.28%          Conservative:   8
Best Local Similarity:    29.17%          Mismatches:     23
Query Match:              12.09%          Indels:         20
DB:                       6               Gaps:           2
```

```
US-09-837-344-19 (1-107) x US-10-131-826A-273 (1-1200)

Qy          4 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeu 23
              :::   ::::::      :::|||   :::   |||||||||          |||   |||
Db        346 GAAGCCCAGCGGGAAGAGCAAAGACTGCGTGTTCACGGAGATCGTGCTGGAGAACAACTA 405

Qy         24 TyrGlyArgLeuGluIleProAlaIleGluLeu-------------------------- 34
              ||||||       |||   ||||||      |||
Db        406 TACGGCCTTCCAGAACGCCCGGCACGAGGGCTGGTTCATGGCCTTCACGCGGCAGGGGCG 465

Qy         35 --------------------ProSerGluAsnGluArgGlyTyrTyr------------ 43
                                  |||   |||      ||||||
Db        466 GCCCCGCCAGGCTTCCCGCAGCCGCCAGAACCAGCGCGAGGCCCACTTCATCAAGCGCCT 525

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArg 55
              :::|||      ::::::|||||||||      |||
Db        526 CTACCAAGGCCAGCTGCCCTTCCCCAACCACGCCGA 561

RESULT 27
US-10-750-185-25422
; Sequence 25422, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 25422
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Bovine  19866881037208
US-10-750-185-25422

Alignment Scores:
Pred. No.:              254           Length:         1400
Score:                  67.00         Matches:        23
Percent Similarity:     35.16%        Conservative:   9
Best Local Similarity:  25.27%        Mismatches:     31
Query Match:            12.09%        Indels:         28
DB:                     6             Gaps:           3

US-09-837-344-19 (1-107) x US-10-750-185-25422 (1-1400)

Qy          6 LysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGly 25
              |||:::   |||||||||
Db        286 AAGCAAACCCTAGAGAGG---------------------------------------- 303

Qy         26 ArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIlePro 45
                    :::|||   :::   ||||||              |||       |||
Db        304 ---------CTGCCTTGCCTCAGTTTACCACCTCATCGAGGGGGAAGCTACATTTTTCCA 354

Qy         46 HisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSer 65
                  ::::::|||      |||||||||      ||||||
Db        355 TCTTCAGCAGCTCTTTCCTCCGACAACAGGAATGAAAGCAGAGGC-------------- 399

Qy         66 IleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIle 85
                    :::   |||      |||:::||||||   |||          ::::::
Db        400 ------CCACAGGCTAACGTAGCATCCTTAACAACCCTAGTTATTGCCTTGATTGAGCTT 453

Qy         86 ------------HisLysGlyHisLeuGluGlu 92
                          |||      |||      |||
Db        454 ATCTAATCAAAGCACAGCCCTCATAGAAGGGAA 486

RESULT 28
US-10-750-185-60116
; Sequence 60116, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 60116
;   LENGTH: 1605
;   TYPE: DNA
;   ORGANISM: Bovine  19866880935399
US-10-750-185-60116

Alignment Scores:
Pred. No.:              304           Length:         1605
Score:                  67.00         Matches:        29
Percent Similarity:     50.00%        Conservative:   12
Best Local Similarity:  35.37%        Mismatches:     33
Query Match:            12.09%        Indels:         8
DB:                     6             Gaps:           5

US-09-837-344-19 (1-107) x US-10-750-185-60116 (1-1605)

Qy         26 ArgLeuGluIleProAlaIleGluLeuProSerGluAsnGlu------ArgGlyTyrTyr 43
              ||||||      |||         |||||||||:::          |||
Db        401 AGATTAACATCACCA------AGACTACCATCAAAGAGATATCCATGCAGAAAGAGGATG 454

Qy         44 IleProHisGlnSerSerLeuProGln---AspAsnArgGlyAsnSerArgAspSerLys 62
              ||||||      |||:::            |||:::||||||   :::||||||   |||
Db        455 ATTCCAAAAACCTCCGCAACCAGCTCGATAGACTCTCGAGGGAGAACCCGAGATCTGAAG 514

Qy         63 GluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArg 82
              |||         |||   :::   |||   :::||||||            |||   |||
Db        515 GAG------GAGATCGTCCGTCTCAAC---GACAGCATCCTGCAGACCACGGAGCAGCGG 565

Qy         83 ArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGln 102
              |||         :::      |||::::::|||      ||||||         ::::::
Db        566 AGGCGGGCCGAGGAGGACGCCCTGCAGCAGAAGGCCTGTGGCTCCGAGATGCTGCAGAAG 625

Qy        103 LysGlu 104
              |||:::
Db        626 AAGCAG 631

RESULT 29
US-10-995-561-13210/c
; Sequence 13210, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13210
```

```
;   LENGTH: 13083
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13210

Alignment Scores:
Pred. No.:              6.33e+03        Length:         13083
Score:                  66.00           Matches:        19
Percent Similarity:     52.73%          Conservative:   10
Best Local Similarity:  34.55%          Mismatches:     18
Query Match:            11.91%          Indels:         8
DB:                     6               Gaps:           2

US-09-837-344-19 (1-107) x US-10-995-561-13210 (1-13083)

Qy         26 ArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIlePro 45
              ||||||   :::|||   ::::::||||||::::::                   |||
Db       9258 AGACTCTTAGTCCCTACCTTGGACCTGCCGAATCAG-----------------ATTCTG 9217

Qy         46 HisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSer 65
              |||         :::|||   |||   ||||||   |||      |||   |||:::
Db       9216 CACTTGCACAGGATCCCAGGTGATGCTCGTGGGCCTTCC------AGTTTGGAAGTGCCC 9163

Qy         66 IleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
              :::      ||||||   ||||||        |||:::   :::
Db       9162 CTGCAAGCTAAAACATACAGGGAGGAGAGGGGGACTAGTTTTGAT 9118

RESULT 30
US-10-995-561-13396/c
; Sequence 13396, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13396
;   LENGTH: 398287
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(398287)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13396

Alignment Scores:
Pred. No.:              5.37e+05        Length:         398287
Score:                  66.00           Matches:        21
Percent Similarity:     47.95%          Conservative:   14
Best Local Similarity:  28.77%          Mismatches:     28
Query Match:            11.91%          Indels:         10
DB:                     6               Gaps:           3

US-09-837-344-19 (1-107) x US-10-995-561-13396 (1-398287)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGlu 21
              |||         ||||||      |||:::||||||   ||||||:::   |||   |||
Db     263051 AAAAAAAAAAAAAAAAAAAAGAAAAGAAAAGAAAAAAAACCATGGTAAT---CTGGAAGAG 262995

Qy         22 AspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGlu------ 39
                 :::         |||   :::      |||:::         :::   |||
Db     262994 AAAATTAAAGAAACTCTTATACTAGCAAAAATTCAAACAGATGGAGATCCTGAGAGAAAA 262935

Qy         40 ---------ArgGlyTyrTyr-----------IleProHisGlnSerSerLeuProGln 52
                      ||||||||                |||      :::   |||   :::
Db     262934 TAAGATTCTAGAGGATAYGCCCAGGAAATCTAAGCCCCATTTAATACTGGTTTGAGAAGA 262875

Qy         53 AspAsnArgGlyAsnSerArgAspSerLysGluIleSer 65
                 :::   :::   |||:::   :::   |||:::|||:::
Db     262874 GAGGCCAAAGACAATGCAGAAGAAATAAAGCAAATAACC 262836

Search completed: December 20, 2005, 05:22:14
Job time : 603 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:38:17 ; Search time 38 Seconds
(without alignments)
270.926 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	553	99.8	1909	2	A45592	liver stage antige
2	91	16.4	210	2	C70309	transcription regu
3	81.5	14.7	442	2	D70315	transcription regu
4	81	14.6	3488	2	T34418	hypothetical prote
5	79.5	14.4	635	1	I37060	involucrin L - gor
6	79.5	14.4	1203	2	S26650	DNA-binding protei
7	79	14.3	511	2	H83339	probable transcrip
8	79	14.3	913	2	T52485	neurofilament prot
9	79	14.3	1125	2	E90598	membrane nuclease,
10	78.5	14.2	615	2	T06108	hypothetical prote
11	78	14.1	772	2	I50463	protein kinase - c
12	77.5	14.0	1706	2	T39305	protein kinase - f
13	77	13.9	550	2	T32560	hypothetical prote
14	77	13.9	762	2	G88436	protein T04A8.13 [
15	77	13.9	789	2	T51310	RNA helicase RH28
16	77	13.9	791	2	T24435	hypothetical prote
17	77	13.9	993	2	B97219	conserved membrane
18	77	13.9	1347	2	F84531	probable retroelem
19	76.5	13.8	199	2	T40870	very hypothetical
20	76.5	13.8	1085	2	S55352	IFH1 protein - yea
21	76	13.7	246	2	F69259	hypothetical prote
22	76	13.7	756	2	T00367	hypothetical prote
23	75.5	13.6	218	2	T16572	hypothetical prote
24	75.5	13.6	765	1	ISHUT1	DNA topoisomerase
25	75.5	13.6	1233	2	S56271	hypothetical prote
26	75	13.5	196	2	D96546	hypothetical prote
27	75	13.5	483	2	S41853	centromere/microtu
28	75	13.5	810	1	S57196	calpain (EC 3.4.22
29	75	13.5	1247	2	E71616	hypothetical prote
30	74.5	13.4	272	2	T26287	hypothetical prote
31	74.5	13.4	473	2	S23901	nitrogen regulatio
32	74.5	13.4	821	2	S67087	hypothetical prote
33	74	13.4	670	2	B70145	periplasmic protei
34	74	13.4	678	2	A54514	glutamic acid-rich
35	74	13.4	816	2	S64439	hypothetical prote
36	74	13.4	1356	2	T16718	hypothetical prote
37	73.5	13.3	2447	2	T16870	hypothetical prote
38	73	13.2	420	2	T44132	plasmid recombinat
39	73	13.2	499	2	S73810	transcription init
40	73	13.2	585	1	A24168	involucrin - human
41	73	13.2	1672	2	T46237	hypothetical prote
42	73	13.2	3600	2	D86161	F10O3.12 protein -
43	72.5	13.1	517	2	E89530	protein H28G03.2 [
44	72.5	13.1	569	2	T39577	hypothetical prote
45	72.5	13.1	605	1	I37061	involucrin M - gor

ALIGNMENTS

RESULT 1
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C;Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R;Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A;Reference number: S24597
A;Accession: S24597
A;Molecule type: DNA
A;Residues: 1-1909 <ZHU>
A;Cross-references: UNIPROT:Q25893; UNIPARC:UPI0000081753; EMBL:X56203; NID:g9915; PID:g
R;Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A;Title: Structure of Plasmodium falciparum liver stage antigen-1.
A;Reference number: A45592; MUID:92107224; PMID:1840628
A;Accession: A45592
A;Molecule type: DNA
A;Residues: 1-195;638-688;1165-1215;1590-1909 <ZH2>
A;Cross-references: UNIPARC:UPI0000177378; UNIPARC:UPI0000177379; UNIPARC:UPI000017737A;
A;Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A;Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A;Reference number: S29393; MUID:87315391; PMID:3306406
A;Accession: S29393
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 323-387 <GUE1>
A;Cross-references: UNIPARC:UPI000017737C; EMBL:M28266
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A;Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by
A;Reference number: S34842
A;Accession: S34842
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 323-381,'HKAI' <GUE2>
A;Cross-references: UNIPARC:UPI000017737D; EMBL:M28266
A;Note: difference at carboxyl end due to frameshift error
C;Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: EF hand
F;154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match 99.8%; Score 553; DB 2; Length 1909;
Best Local Similarity 99.1%; Pred. No. 3.5e-43;
Matches 106; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

```
Db       1630 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 1689

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1690 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 1736


RESULT 2
C70309
transcription regulator - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: C70309
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: C70309
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-210 <AQF>
A;Cross-references: UNIPROT:O66502; UNIPARC:UPI0000056271; GB:AE000673; NID:g2982834; PI
A;Experimental source: strain VF5
C;Genetics:
A;Gene: hth

  Query Match             16.4%;  Score 91;  DB 2;  Length 210;
  Best Local Similarity   28.4%;  Pred. No. 0.28;
  Matches   40;  Conservative   18;  Mismatches   37;  Indels   46;  Gaps    7;

Qy          1 RKAD------TKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSS----- 49
              ||||      | |||::  | |    |||| || :  :|||   ||   |||  |
Db         41 RKADVRIIAATNKNLKKLVEEGK-FREDLYYRLNVVNVELPPLRERKEDIPHLVSYFLEK 99

Qy         50 --------------------LPQDNRGNSRDSKEISIIEKTNRESI---TTNVEGRRDIH 86
                                  | |  || |:      :|    ::   :|:: |  |:
Db        100 FSKIHGKKVKGVSPEAMKLLLEYDYPGNVRE------LENAIEHAVVVSSTSLIGPDDLP 153

Qy         87 KGHLEEKKDGS-IKPEQKEDK 106
                 || : || :| : :|::
Db        154 ----EEIRIGSKVKKKFRENE 170


RESULT 3
D70315
transcription regulator NtrC family - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2003
C;Accession: D70315
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: D70315
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-442 <AQF>
A;Cross-references: UNIPARC:UPI00000562A2; GB:AE000676; NID:g2982884; PIDN:AAC06509.1; P
A;Experimental source: strain VF5
C;Genetics:
A;Gene: ntrC4
C;Superfamily: response regulator, NtrC type; response regulator homology; RNA polymeras
C;Keywords: phosphoprotein
F;4-113/Domain: response regulator homology <RRH>
F;132-353/Domain: RNA polymerase sigma factor interaction domain homology <SFI>
F;52/Binding site: phosphate (Asp) (covalent) #status predicted

  Query Match             14.7%;  Score 81.5;  DB 2;  Length 442;
  Best Local Similarity   24.1%;  Pred. No. 5.1;
  Matches   33;  Conservative   20;  Mismatches   35;  Indels   49;  Gaps    5;

Qy          5 TKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERG--------YYI------------ 44
              | |||| : : |:   |||| || :  | ||   |||         |::
Db        271 TNKNLEEEIKKGN-FREDLYYRLSVFQIYLPPLRERGKDVILLAEYFLKKFAKEYKKNCF 329

Qy         45 -----PHQSSLPQDNRGNSRDSK-----------------EISIIEKTNRE-----SIT 76
                  :  : |: :|| |: |                   :: : ||: |:      |
Db        330 ELSEETKEYLMKQEWKGNVRELKNLIERAVILCEGEVIKPEDLGLKEKSWRDLSYLLKIK 389

Qy         77 TNVEGRRDIHKGHLEEK 93
                | :::  |  :|||
Db        390 ELKEAKKEFEKIFIEEK 406


RESULT 4
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: UNIPARC:UPI000017B8E6; EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023;
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

  Query Match             14.6%;  Score 81;  DB 2;  Length 3488;
  Best Local Similarity   23.3%;  Pred. No. 64;
  Matches   31;  Conservative   19;  Mismatches   47;  Indels   36;  Gaps    3;

Qy          2 KADTKKNLERKKEHGD--------------------ILAEDLYGRLEIPAIELPSENERG 41
              |  ||| : :|||  :                    :||   |  : |  | |::
Db        361 KPKTKKKIIKKKETPESEQVTAAEPEQQKISEVDVQSVAETEVGAKKKPDAEKPTD---- 416

Qy         42 YYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKP- 100
                 |   :|::    |  | |  ||:  |  | :   ::   |  :: ||: : ||
Db        417 -----LSKAKKDSKSKKSDEPEASTEEKSTTEKPTNDKTSKKSAEKKTVKPKKEVTGKPL 471

Qy        101 ------EQKEDKS 107
                   | |:| |
Db        472 EAKKPVEDKKDAS 484


RESULT 5
I37060
involucrin L - gorilla
C;Species: Gorilla gorilla (gorilla)
C;Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C;Accession: I37060
R;Teumer, J.; Green, H.
Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286, 1989
A;Title: Divergent evolution of part of the involucrin gene in the hominoids: Unique int
A;Reference number: I37060; MUID:89145226; PMID:2919176
A;Accession: I37060
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-635 <RES>
A;Cross-references: UNIPROT:Q28423; UNIPARC:UPI0000087628; GB:M23603; GB:J04499; NID:g34
C;Comment: During the terminal differentiation of keratinocytes, this protein from the c
linked envelope under the plasma membrane.
C;Superfamily: involucrin
C;Keywords: cornified cell envelope; duplication; epidermis; tandem repeat
F;153-590/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

  Query Match             14.4%;  Score 79.5;  DB 1;  Length 635;
```

```
  Best Local Similarity   26.8%;  Pred. No. 12;
  Matches   33;  Conservative   24;  Mismatches   35;  Indels   31;  Gaps    8;

Qy          7 KNLERKKEHGDI-LAEDLYGRLEIP---------------AIELPSENERGYYIPHQS-- 48
              |:||::|  | : | |   |:||:|                 :||| : |    :| |
Db        164 KHLEQQK--GQLELPEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQQEG 221

Qy         49 --SLPQDNRGNSR--DSKEISIIEKTNRESITTNVEGRRDIHKGHLE--EKKDGSIK-PE 101
                ||:   |  :  : :|  :     :|    ::| :    || ||  |:::| :: ||
Db        222 QLELPEQQEGQLKHLEQQEGQLELPEQQEGQLKHLEQQ----KGQLELPEQQEGQLELPE 277

Qy        102 QKE 104
              |:|
Db        278 QQE 280


RESULT 6
S26650
DNA-binding protein 5 - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S26650; PN0099; I57397
R;Mattioni, T.; Hume, C.R.; Konigorski, S.; Hayes, P.; Osterweil, Z.; Lee, J.S.
Chromosoma 101, 618-624, 1992
A;Title: A cDNA clone for a novel nuclear protein with DNA binding activity.
A;Reference number: S26650; MUID:93048367; PMID:1424986
A;Accession: S26650
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1203 <MAT>
A;Cross-references: UNIPARC:UPI000017C0E3; EMBL:X63071
R;Berdichevsky, F.B.; Chumakov, I.M.; Kisselev, L.L.
Mol. Biol. (Mosk.) 22, 794-801, 1988
A;Title: Determination of the nucleotide sequence of the human genome son3 fragment: ide
A;Reference number: PN0099; MUID:89039788; PMID:3054499
A;Accession: PN0099
A;Molecule type: mRNA
A;Residues: 'R',574-699,'PH',702-1055 <BER>
A;Cross-references: UNIPARC:UPI000016B047
A;Experimental source: embryo
A;Note: this paper is in Russian, with an English abstract
R;Berdichevsky, F.B.; Chumakov, I.M.; Kiselev, L.L.
Mol. Biol. 22, 639-646, 1988
A;Title: Determination of the nucleotide sequence of the son3 fragment of the human genc
A;Reference number: I57397
A;Accession: I57397
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'R',574-699,'PH',702-1055 <RES>
A;Cross-references: UNIPARC:UPI000016B047; GB:M36428; NID:g338289; PIDN:AAA36624.1; PID:
C;Genetics:
A;Introns: 790/3
C;Keywords: DNA binding
F;741-928/Region: arginine/serine-rich repeats

  Query Match             14.4%;  Score 79.5;  DB 2;  Length 1203;
  Best Local Similarity   26.6%;  Pred. No. 25;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   25;  Gaps    5;

Qy          7 KNLERKKE-----HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQ---DNRGNS 58
              |:||          | :|| |: ||             :|     |  ||:|:   ::
Db        620 KDMERLTSLRAGIEGPLLASDV-GR-----------DRSAASPVVSSMPERASESSSEE 666

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              :|  || :  |   |     |    ::  ||  |:|:| |::   |  ||
Db        667 KDDYEIFVKVKDTHEKSKKN----KNRDKGEKEKKRDSSLRSRSKRSKS 711


RESULT 7
H83339
probable transcription regulator PA2449 [imported] - Pseudomonas aeruginosa (strain PAO1
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: H83339
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83339
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-511 <STO>
A;Cross-references: UNIPROT:Q9I133; UNIPARC:UPI00000C56DB; GB:AE004672; GB:AE004091; NI
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2449
C;Superfamily: transcriptional regulator of TyrR type with ACT, PAS, AAA, and Fis domai

  Query Match             14.3%;  Score 79;  DB 2;  Length 511;
  Best Local Similarity   26.8%;  Pred. No. 10;
  Matches   34;  Conservative   16;  Mismatches   41;  Indels   36;  Gaps    5;

Qy          5 TKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI----PH--QSSLPQDNR--- 55
              | :|||:    |    |||: || : ::|:|   |||: |      |  | :  |  |
Db        343 THRNLEKMVAEGS-FREDLFYRLNVLSLEVPPLRERGHDILLLARHFMQQACAQIQRPVC 401

Qy         56 ----------------GNSRDSK----------EISIIEKTNRESITTNVEGRRDIHKGH 89
                             || |  :          | |:::  : |   | |  : |   |
Db        402 RLAPGTYPALLSNRWPGNVRQLQNVIFRAAAICESSLVDIGDLEIAGTAVARQNDGEVGS 461

Qy         90 LEEKKDG 96
              |||  :|
Db        462 LEEAVEG 468


RESULT 8
T52485
neurofilament protein NF-M(2), middle molecular weight [imported] - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 09-Jul-2004
C;Accession: T52485
R;Gervasi, C.; Szaro, B.G.
Brain Res. Mol. Brain Res. 48, 229-242, 1997
A;Title: Sequence and expression pattern of two forms of the middle molecular weight ne
A;Reference number: Z26090
A;Accession: T52485
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-913 <GER>
A;Cross-references: UNIPROT:O13099; UNIPARC:UPI00000FCFAD; EMBL:U85970; PIDN:AAB53390.1

  Query Match             14.3%;  Score 79;  DB 2;  Length 913;
  Best Local Similarity   28.7%;  Pred. No. 20;
  Matches   29;  Conservative   16;  Mismatches   44;  Indels   12;  Gaps    4;

Qy          4 DTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKE 63
              | |   | :|| |:  ||       |  | |   | |:|     :     ::  |   : :|
Db        562 DVKAEEEEEKEEGEEDAE----VEETKAEEGEEEGEKG-----EDEAEKEEGGEEEEKEE 612

Qy         64 ISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKE 104
              ::  |:   |      || :|  ||  ||:|:|  : |: |
Db        613 VAADEE-GGEGEEEGGEGEKDEDKG--EEEKEGEAEAEEAE 650


RESULT 9
E90598
membrane nuclease, lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: E90598
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A;Reference number: A99512; MUID:21267165; PMID:11353084
```

```
A;Accession: E90598
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1125 <KUR>
A;Cross-references: UNIPROT:Q98PM9; UNIPARC:UPI00000D464B; GB:AL445566; PID:g14090108; P
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_6930
A;Genetic code: SGC3

  Query Match             14.3%;  Score 79;  DB 2;  Length 1125;
  Best Local Similarity   26.4%;  Pred. No. 26;
  Matches   28;  Conservative   16;  Mismatches   36;  Indels   26;  Gaps    4;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              :| |   || | | |: :   ::|         : |  |           : |:    |
Db        466 QKDDQNKNSENKDENSNDSKQNL---------DKPKTNN-----------EQNQNTQDD 504

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              ||:||   : |  | |||    :|      :| |: :|| :| : |
Db        505 SKKISDASQNN--SNTTN---EKDQKLDSQDESKNNAIKSQQNDQK 545


RESULT 10
T06108
hypothetical protein T5J17.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C;Accession: T06108
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z15184
A;Accession: T06108
A;Molecule type: DNA
A;Residues: 1-615 <BEV>
A;Cross-references: UNIPROT:Q9SMQ3; UNIPARC:UPI00000AB01D; EMBL:AL035708; GSPDB:GN00062;
A;Experimental source: cultivar Columbia; BAC clone T5J17
C;Genetics:
A;Gene: ATSP:T5J17.190
A;Map position: 4

  Query Match             14.2%;  Score 78.5;  DB 2;  Length 615;
  Best Local Similarity   29.4%;  Pred. No. 14;
  Matches   30;  Conservative   20;  Mismatches   41;  Indels   11;  Gaps    4;

Qy          7 KNLERKKEHGDILAEDLYGRLE--IPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEI 64
              | :|| | : : :| |   :|:  :  ||:  | |:   :  | |:|::      :  |
Db        342 KEIERVKVN-EAVANDNIKKLKKMLSEIEVAMEEEKQRSLNRQESMPKE----VVEVVEK 396

Qy         65 SIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              | ||  :|    | : :::  |    |||: | | | || |
Db        397 KIEEKEKKEEKKENKKEKKESKK----EKKEHSEKKEDKEKK 434


RESULT 11
I50463
protein kinase - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 05-Oct-2004
C;Accession: I50463
R;Li, H.; Grenet, J.; Valentine, M.; Lahti, J.M.; Kidd, V.J.
Gene 153, 237-242, 1995
A;Title: Structure and expression of chicken protein kinase PITSLRE-encoding genes.
A;Reference number: I50463; MUID:95180725; PMID:7875595
A;Accession: I50463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-772 <LIX>
A;Cross-references: UNIPROT:Q91013; UNIPARC:UPI00000FC347; EMBL:U16656; NID:g571458; PID
C;Genetics:
A;Gene: cPITSLRE
A;Introns: 37/3; 76/2; 118/1; 164/2; 208/1; 249/1; 283/1; 329/1; 351/1; 407/3; 438/1; 47
C;Keywords: ATP
F;413-668/Domain: protein kinase homology <KIN>
F;421-429/Region: protein kinase ATP-binding motif

  Query Match             14.1%;  Score 78;  DB 2;  Length 772;
  Best Local Similarity   27.0%;  Pred. No. 21;
  Matches   30;  Conservative   18;  Mismatches   45;  Indels   18;  Gaps    5;

Qy          2 KADTK--KNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSR 59
              ||: |  || : :    | | |   | |    :|:   |       |::     ::||
Db         29 KAEIKRMKNSDDRDSKRDSLEE---GELRDHRMEITIRNS-----PYRREDSMEDRGEED 80

Qy         60 DS------KEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKE 104
              ||      :::|  |||:        | ||  |: |  | |   :| :::|
Db         81 DSLAIKPPQQMSRKEKTHHRKDEKRKEKRR--HRSHSAEGKHARVKEKERE 129


RESULT 12
T39305
protein kinase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T39305
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z21843
A;Accession: T39305
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1706 <WOO>
A;Cross-references: UNIPROT:O42900; UNIPARC:UPI000006988D; EMBL:AL022117; PIDN:CAA17922.
A;Experimental source: strain 972h-; cosmid c119
C;Genetics:
A;Gene: SPDB:SPBC119.07
A;Map position: 2
A;Introns: 29/2; 72/1; 264/3

  Query Match             14.0%;  Score 77.5;  DB 2;  Length 1706;
  Best Local Similarity   25.3%;  Pred. No. 59;
  Matches   25;  Conservative   20;  Mismatches   47;  Indels    7;  Gaps    4;

Qy          8 NLERKKEHGDIL--AEDLYGRLEIPAIELPSENERGYYIPHQSSL--PQDNRG--NSRDS 61
              | |   ::|: :    |:| : : | |:|||:        |  ::  | |: |  || ::
Db        998 NRESNSDNGETITGTHDVYRQTDNPEIKLPSDTASSKVDTHNPTVTQPTDDTGGLNSYNT 1057

Qy         62 KEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKP 100
              :   :   |   |    :   :|  | | :| |  |: |
Db       1058 ENPLLTNNTLEPSSVEAIVSSKDSDK-HAKESKGKSLAP 1095


RESULT 13
T32560
hypothetical protein T11F8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T32560
R;Johnson, D.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid T11F8.
A;Reference number: Z21191
A;Accession: T32560
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-550 <JOH>
A;Cross-references: UNIPROT:O44192; UNIPARC:UPI000016419C; EMBL:AF036703; PIDN:AAB88557.
A;Experimental source: strain Bristol N2; clone T11F8
C;Genetics:
A;Gene: CESP:T11F8.4
A;Map position: 4
A;Introns: 16/3; 147/3; 201/3; 268/2; 387/3

  Query Match             13.9%;  Score 77;  DB 2;  Length 550;
  Best Local Similarity   22.8%;  Pred. No. 17;
  Matches   26;  Conservative   21;  Mismatches   35;  Indels   32;  Gaps    3;
```

```
Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              :::| :|| :|||:                   |   :  :|  |  :|   :|:| : |
Db        411 QESDKRKNSDRKKKS----------------ERKKKGSKGDMILEKSKETKDHRESGR- 452

Qy         61 SKEISIIEKTNRE-------SITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
                        |:||       ||    :| :|: :  :    | ||| |: |  ::
Db        453 -------RLTDREANAVGGKSIDAKSKGNKDMEQRTIMIDKPGSITPKSKRSRT 499


RESULT 14
G88436
protein T04A8.13 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: G88436
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: G88436
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-762 <STO>
A;Cross-references: UNIPROT:Q22142; UNIPARC:UPI000017A5C4; GB:chr_III; PIDN:CAA84732.1;
C;Genetics:
A;Gene: T04A8.13
A;Map position: 3

  Query Match             13.9%;  Score 77;  DB 2;  Length 762;
  Best Local Similarity   28.6%;  Pred. No. 25;
  Matches   30;  Conservative   12;  Mismatches   43;  Indels   20;  Gaps    3;

Qy          2 KADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDS 61
              | : :|  | ||| |    ||          |   |||             : :  |:|
Db         75 KNEKEKKEEEKKEDGHEKKED--------KKEDKKENEN----------DEKKEKSKDD 115

Qy         62 KEISIIEKTNRESITTNVEGRRDIHKGH-LEEKKDGSIKPEQKED 105
              |:    |    :: | : ||: :  |    |||||   | |: ||
Db        116 KKEESKEDKKEKTKTEDNEGKNEEKKKEKKEEKKDEKEKKEKTED 160


RESULT 15
T51310
RNA helicase RH28 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004
C;Accession: T51310
R;Aubourg, S.; Kreis, M.; Lecharny, A.
Nucleic Acids Res. 27, 628-636, 1999
A;Title: The DEAD box RNA helicase family in Arabidopsis thaliana.
A;Reference number: Z22965; MUID:9862990; PMID:9862990
A;Accession: T51310
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-789 <AUB>
A;Cross-references: UNIPROT:Q9ZRZ8; UNIPARC:UPI00000A9135; EMBL:AJ010475; PIDN:CAA09214.
C;Genetics:
A;Gene: RH28

  Query Match             13.9%;  Score 77;  DB 2;  Length 789;
  Best Local Similarity   26.8%;  Pred. No. 26;
  Matches   33;  Conservative   18;  Mismatches   30;  Indels   42;  Gaps    7;

Qy          1 RKADTKKNLERKK----EHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRG 56
              || : :||| |||     |    : ||  | |         |:| |           : ||
Db        641 RKRERENKNLPRKKRRKLEAAREMLEDNEGEEE-------EEDEEG----------DEKRG 683

Qy         57 NSR--DSKEISIIEKTNRESITTNVEG----------RRDIHKGHLEEKKDGSIKPEQKE 104
              ||   | |:    ::|::: :|       :| |  | :|       ||::|:
Db        684 RSRGKDKKK----QETDKKGLTLKDLGYMRAKAVKAKQRAIDSGKMERP-----KPDKKQ 734

Qy        105 DKS 107
                :|
Db        735 SRS 737


RESULT 16
T24435
hypothetical protein T04A8.13 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24435
R;Palmer, S.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19889
A;Accession: T24435
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-791 <WIL>
A;Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PIDN:CAA84732.2
A;Experimental source: clone T04A8
C;Genetics:
A;Gene: CESP:T04A8.13
A;Map position: 3
A;Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

  Query Match             13.9%;  Score 77;  DB 2;  Length 791;
  Best Local Similarity   28.6%;  Pred. No. 26;
  Matches   30;  Conservative   12;  Mismatches   43;  Indels   20;  Gaps    3;

Qy          2 KADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDS 61
              | : :|  | ||| |    ||          |   |||             : :  |:|
Db         75 KNEKEKKEEEKKEDGHEKKED--------KKEDKKENEN----------DEKKEKSKDD 115

Qy         62 KEISIIEKTNRESITTNVEGRRDIHKGH-LEEKKDGSIKPEQKED 105
              |:    |    :: | : ||: :  |    |||||   | |: ||
Db        116 KKEESKEDKKEKTKTEDNEGKNEEKKKEKKEEKKDEKEKKEKTED 160


RESULT 17
B97219
conserved membrane protein CAC2592 (6-pyruvoyl-tetrahydropterin synthase related domain)
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: B97219
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97219
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-993 <KUR>
A;Cross-references: UNIPROT:Q97FY2; UNIPARC:UPI00000CA59A; GB:AE001437; PIDN:AAK80541.1,
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2592

  Query Match             13.9%;  Score 77;  DB 2;  Length 993;
  Best Local Similarity   28.3%;  Pred. No. 35;
  Matches   30;  Conservative   23;  Mismatches   37;  Indels   16;  Gaps    6;

Qy         16 GDILAEDLYGRLEIPAI-----ELPSENERGYYIPH--QSSLPQDN---RGNSRDSKEI- 64
              ||    ::|  ||::  |      |:       | | : |  :: : :|:   :|    : | :
Db         84 GDFFFDELKDRLKLKGIFLNKLEISETPARVYVVNHHRENVVFEDSKKKKGYDNNIKTLI 143

Qy         65 --SIIEKTNRESITTNVE--GRRDIHKGHLEEKKDGSIKPEQKEDK 106
                |:|  |  | |  |||    |: :||  ||| :: :|| :: : |
Db        144 MKSLINSTTDEIIKDNVEILKRKIVHK-ELEETREKTIKRDRTKLK 188


RESULT 18
```

```
F84531
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: F84531
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1347 <STO>
A;Cross-references: UNIPROT:Q9ZQE9; UNIPARC:UPI000009EC74; GB:AE002093; NID:g4335731; PI
C;Genetics:
A;Gene: At2g15650
A;Map position: 2
C;Superfamily: retrovirus-related polyprotein

  Query Match             13.9%;  Score 77;  DB 2;  Length 1347;
  Best Local Similarity   26.5%;  Pred. No. 49;
  Matches   31;  Conservative   16;  Mismatches   34;  Indels   36;  Gaps    6;

Qy         16 GDILAEDLYGR-----------LEIPAIELPSENERGYYIPHQSS--LPQDNRGNSRDSK 62
                 |||     :|:           ||   : :|     ||::  |    : ||  |     |
Db        368 GDITVMTRHGKRIIKNVFLVPGLEKNLLSVPQIISSGYWVRFQDKRCIIQDANG-----K 422

Qy         63 EISIIEKTNR-----------ESITTNVEGRRDIHK--GHLEEKKDGSIKPEQKEDK 106
                 ||  || |::          |::| ||:     ||  ||:  |     : :| :||
Db        423 EIMNIEMTDKSFKIKLSSVEEEAMTANVQTEETWHKRLGHVSNK-----RLQQMQDK 474


RESULT 19
T40870
very hypothetical protein SPCC1223.10c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40870
R;Wedler, H.; Wambutt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21953
A;Accession: T40870
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-199 <WED>
A;Cross-references: UNIPROT:O74410; UNIPARC:UPI000006B71C; EMBL:AL031579; PIDN:CAA20879.
A;Experimental source: strain 972h-; cosmid c1223
C;Genetics:
A;Gene: SPDB:SPCC1223.10c
A;Map position: 3
A;Introns: 7/3; 36/3

  Query Match             13.8%;  Score 76.5;  DB 2;  Length 199;
  Best Local Similarity   30.6%;  Pred. No. 5.8;
  Matches   19;  Conservative   10;  Mismatches   30;  Indels    3;  Gaps    1;

Qy         45 PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKE 104
              | |     |     |: :||  : |:: ::|||   |::    ||  ||| |||
Db         55 PKVSKTVPSNAITQSDNSQISESKSTSQSAVTTNSTRRKE---KELEASKDGKIKPSSSN 111

Qy        105 DK 106
               :
Db        112 TR 113


RESULT 20
S55352
IFH1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein L8083.9; protein YLR223c; RRP3 protein
C;Species: Saccharomyces cerevisiae
C;Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S55352; S51446; S47477
R;Cherel, I.; Thuriaux, P.
Yeast 11, 261-270, 1995
A;Title: The IFH1 gene product interacts with a fork head protein in Saccharomyces cerev
A;Reference number: S55352; MUID:95304839; PMID:7785326
A;Accession: S55352
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1085 <CHE>
A;Cross-references: UNIPROT:P39520; UNIPARC:UPI000012D3D6; EMBL:Z29488; NID:g531491; PID
R;Hallsworth, K.
submitted to the EMBL Data Library, December 1994
A;Description: The sequence of S. cerevisiae cosmid 8083.
A;Reference number: S51443
A;Accession: S51446
A;Molecule type: DNA
A;Residues: 1-1085 <HAL>
A;Cross-references: UNIPARC:UPI000012D3D6; EMBL:U19027; NID:g609363; PID:g609372; MIPS:Y
C;Genetics:
A;Gene: SGD:IFH1; RRP3
A;Cross-references: SGD:S0004213; MIPS:YLR223c
A;Map position: 12R

  Query Match             13.8%;  Score 76.5;  DB 2;  Length 1085;
  Best Local Similarity   21.3%;  Pred. No. 43;
  Matches   30;  Conservative   15;  Mismatches   53;  Indels   43;  Gaps    4;

Qy          1 RKADTKKNLERKKE----------------------------------HGDI-LAEDLY 24
              :| : |: |::||:                                  | ||    |
Db        873 KKTEKKRQLKKKKKLLKIRKQRQKAIKEQETMNLQLGINGHEIIGNNNSHSDINTGTDFT 932

Qy         25 GRLEIPAIELPSE-NERGYYIPHQSSLPQDNRGNSRDSKEI------SIIEKTNRESITT 77
                 |  ||||   |    ||| | |  |:       :| :       |: |   : ::
Db        933 TNENTPMNELPSHAPEDASLIPHNSDLAVDSNTRKNSTKSVGLDEIHEILGKDENDLLSV 992

Qy         78 NVEGRRDIHKGHLEEKKDGSI 98
                    |  :||: |  |  |
Db        993 GDINGYDAQEGHVIEDTDADI 1013


RESULT 21
F69259
hypothetical protein AF0078 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: F69259
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: F69259
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-246 <KLE>
A;Cross-references: UNIPROT:O30158; UNIPARC:UPI0000057249; GB:AE001101; GB:AE000782; NID
C;Superfamily: Archaeoglobus fulgidus hypothetical protein AF0078

  Query Match             13.7%;  Score 76;  DB 2;  Length 246;
  Best Local Similarity   28.3%;  Pred. No. 8.3;
  Matches   28;  Conservative   12;  Mismatches   39;  Indels   20;  Gaps    3;

Qy         18 ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-------------PQDNRGNSRDSKE 63
              : ||  |  | |  : |    ||| |:  | ||               :: |   || :|
Db         27 VQAETKYSLLRIEYLSL--LEERGKYVEQQKSLAERLQNCTKMLEKAAEERRDVERDVEE 84

Qy         64 ISIIEKTNRESITTNVEG----RRDIHKGHLEEKKDGSI 98
              :  : ||  |  |  :||    | |  |  :|:  :  |
Db         85 LKDLISTNPERWTVRLEGSIVERLDKAKRAIEDSLNSPI 123
```

```
RESULT 22
T00367
hypothetical protein KIAA0665 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C;Accession: T00367
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A;Reference number: Z14142; MUID:98403880; PMID:9734811
A;Accession: T00367
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-756 <ISH>
A;Cross-references: UNIPROT:O75154; UNIPARC:UPI0000129C9C; EMBL:AB014565; NID:g3327143;
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0665

  Query Match             13.7%;  Score 76;  DB 2;  Length 756;
  Best Local Similarity   22.8%;  Pred. No. 31;
  Matches   28;  Conservative   22;  Mismatches   49;  Indels   24;  Gaps    4;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLE-IPAIELPSENERG---------YYIPHQSS- 49
              |:  :   :| ||     | |||  :|| :  ::| :|  ||          |:   : |
Db        603 RLSHERHQFQRDKEATQELIEDLRKQLEHLQLLKLEAEQRRGRSSSMGLQEYHSRARESE 662

Qy         50 -------LPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ 102
                    | ||||     ::|:      | : || :::| : :      |      |
Db        663 LEQEVRRLKQDNRNLKEQNEEL------NGQIITLSIQGAKSLPSTAFSESLAAEISSVS 716

Qy        103 KED 105
              :::
Db        717 RDE 719


RESULT 23
T16572
hypothetical protein K05F1.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T16572
R;Wohldmann, P.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid K05F1.
A;Reference number: Z18537
A;Accession: T16572
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-218 <WOH>
A;Cross-references: UNIPROT:Q21246; UNIPARC:UPI00000825F9; EMBL:U29377; NID:g868173; PID
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:K05F1.9
A;Introns: 25/3; 82/3; 112/3

  Query Match             13.6%;  Score 75.5;  DB 2;  Length 218;
  Best Local Similarity   26.6%;  Pred. No. 8;
  Matches   21;  Conservative   17;  Mismatches   30;  Indels   11;  Gaps    2;

Qy         28 EIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHK 87
              |:|    |:| | |        :  | :   :| |:     | : :   :: ||::|  |
Db        140 ELP----PTEAEDG-------KMGDDKKKGKKDDKKEEKDGKKDEKKDGSDEEGKKDEGK 188

Qy         88 GHLEEKKDGSIKPEQKEDK 106
              |  :|:|    | ::|:||
Db        189 GEKKEEKKDDKKDDKKDDK 207


RESULT 24
ISHUT1
DNA topoisomerase (EC 5.99.1.2) - human
N;Alternate names: nicking-closing enzyme; omega-protein; relaxing enzyme; swivelase; ty
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C;Accession: A30887; A40008; S13821; S02397; S40643; A34422
R;D'Arpa, P.; Machlin, P.S.; Ratrie III, H.; Rothfield, N.F.; Cleveland, D.W.; Earnshaw,
Proc. Natl. Acad. Sci. U.S.A. 85, 2543-2547, 1988
A;Title: cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67.7-kDa car
A;Reference number: A30887; MUID:88190108; PMID:2833744
A;Accession: A30887
A;Molecule type: mRNA
A;Residues: 1-765 <DAR>
A;Cross-references: UNIPROT:P11387; UNIPARC:UPI00001421A9; GB:J03250; NID:g339805; PIDN
R;Kunze, N.; Yang, G.; Doelberg, M.; Sundarp, R.; Knippers, R.; Richter, A.
J. Biol. Chem. 266, 9610-9616, 1991
A;Title: Structure of the human type I DNA topoisomerase gene.
A;Reference number: A40008; MUID:91236733; PMID:1851751
A;Accession: A40008
A;Molecule type: DNA
A;Residues: 1-144,'A',146-553,'E',555-765 <KUN>
A;Cross-references: UNIPARC:UPI00001730AD; GB:M60688; GB:M60689; GB:M60690; GB:M60691; G
:M60702; GB:M60703; GB:M60704; GB:M60705; GB:M60706
R;Kunze, N.; Klein, M.; Richter, A.; Knippers, R.
Eur. J. Biochem. 194, 323-330, 1990
A;Title: Structural characterization of the human DNA topoisomerase I gene promoter.
A;Reference number: S13821; MUID:91099302; PMID:2176592
A;Accession: S13821
A;Molecule type: DNA
A;Residues: 1-20 <KU2>
A;Cross-references: UNIPARC:UPI00001730AE; EMBL:X52601
R;Oddou, P.; Schmidt, U.; Knippers, R.; Richter, A.
Eur. J. Biochem. 177, 523-529, 1988
A;Title: Monoclonal antibodies neutralizing mammalian DNA topoisomerase I activity.
A;Reference number: S02397; MUID:89064806; PMID:2461859
A;Accession: S02397
A;Molecule type: mRNA
A;Residues: 344-765 <ODD>
A;Cross-references: UNIPARC:UPI00001730AF; GB:M60657
R;Tamura, H.O.; Kohchi, C.; Yamada, R.; Ikeda, T.; Koiwai, O.; Patterson, E.; Keene, J.D
Nucleic Acids Res. 19, 69-75, 1991
A;Title: Molecular cloning of a cDNA of a camptothecin-resistant human DNA topoisomerase
A;Reference number: S40643; MUID:91187651; PMID:1849260
A;Accession: S40643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 523-543;573-582,'D',584-593 <TAM>
A;Cross-references: UNIPARC:UPI00001730B0; UNIPARC:UPI00001730B1
R;Maul, G.G.; Jimenez, S.A.; Riggs, E.; Ziemnicka-Kotula, D.
Proc. Natl. Acad. Sci. U.S.A. 86, 8492-8496, 1989
A;Title: Determination of an epitope of the diffuse systemic sclerosis marker antigen DN
ity in systemic sclerosis.
A;Reference number: A34422; MUID:90046823; PMID:2479024
A;Accession: A34422
A;Molecule type: mRNA
A;Residues: 657-765 <MAU>
A;Cross-references: UNIPARC:UPI000016B0FE; GB:M27913; NID:g339807; PIDN:AAA61208.1; PID:
C;Comment: Type I DNA topoisomerase catalyzes the ATP-independent transient breakage of
 in another, followed by rejoining. This reaction will lead to the conversion of one top
C;Genetics:
A;Gene: GDB:TOP1
A;Cross-references: GDB:120444; OMIM:126420
A;Map position: 20q12-20q13.1
A;Introns: 11/3; 20/1
A;Note: the list of introns is incomplete
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: DNA binding; DNA replication; heterotetramer; isomerase
F;723/Active site: Tyr #status predicted

  Query Match             13.6%;  Score 75.5;  DB 1;  Length 765;
  Best Local Similarity   26.9%;  Pred. No. 35;
  Matches   29;  Conservative   18;  Mismatches   40;  Indels   21;  Gaps    5;

Qy          6 KKNLERKKEHGD--ILAEDLYGRLEIPAIELPSENERGYYIPHQSSL-----PQDNRGNS 58
              |:  |: :  ||  |  |   |    | |:   |:: ||::| :  :     |:|
```

```
Db         88 KRKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDD-GYFVPPKEDIKPLKRPRDEDDVD 146

Qy         59 RDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                 |:|   | | :|           |  |||::|| :|  : :||
Db        147 YKPKKIK-TEDTKKEK-----------KRKLEEEEDGKLKKPKNKDK 181


RESULT 25
S56271
hypothetical protein YFR016c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C;Accession: S56271
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanu
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccaromyces ce
A;Reference number: S56186
A;Accession: S56271
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1233 <MUR>
A;Cross-references: UNIPROT:P43597; UNIPARC:UPI000013AEA4; EMBL:D50617; NID:g836685; PID
C;Genetics:
A;Cross-references: SGD:S0001912
A;Map position: 6R

  Query Match             13.6%;  Score 75.5;  DB 2;  Length 1233;
  Best Local Similarity   29.6%;  Pred. No. 61;
  Matches   32;  Conservative   21;  Mismatches   36;  Indels   19;  Gaps    7;

Qy          2 KADTK--KNLERKKEHGDILAE---DLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRG 56
              :| ||  :|:|  |     :||   |: |  ||  | :  | :|   :|      || :
Db        942 EAVTKEDENMENSK-----IAEALKDVTGDQEIDDINISDEFQRTVELPELEK--QDIKD 994

Qy         57 NSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKE 104
              |  : ||:  :|:| :|:   ::    :|     ||| :  || |::|
Db        995 NKGEDKELE-VEETEKETSLPDLVVEENI----TEEKNE--IKQEEEE 1035


RESULT 26
D96546
hypothetical protein F8A12.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D96546
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96546
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <STO>
A;Cross-references: UNIPROT:Q9C6I7; UNIPARC:UPI00000A9647; GB:AE005173; NID:g11094696; P
C;Genetics:
A;Gene: F8A12.15
A;Map position: 1

  Query Match             13.5%;  Score 75;  DB 2;  Length 196;
  Best Local Similarity   26.1%;  Pred. No. 7.9;
  Matches   24;  Conservative   19;  Mismatches   33;  Indels   16;  Gaps    3;

Qy         14 EHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRE 73
              :| | : |              :| |  :|  : ||:  |       :| |: : :
Db         41 DHDDEMVE-------------TEGEMTHYYDNDSSMISDAASPVHTTK-INNVVRRKAN 85

Qy         74 SITTNVEGRRDIHKGHLEEKKDGSIKPEQKED 105
              :| || : || ||: | ||:::   | |::|:
Db         86 NINTNPKKRRIIHQ-HKEEEEEELQKGEEEEE 116


RESULT 27
S41853
centromere/microtubule-binding protein CBF5 [validated] - yeast (Saccharomyces cerevisia
N;Alternate names: protein L9470.11; protein YLR175w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change 09-Jul-2004
C;Accession: S41853; S45443; S51420
R;Jiang, W.; Middleton, K.; Yoon, H.J.; Fouquet, C.; Carbon, J.
Mol. Cell. Biol. 13, 4884-4893, 1993
A;Title: An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubul
A;Reference number: S41853; MUID:93330283; PMID:8336724
A;Accession: S41853
A;Molecule type: DNA
A;Residues: 1-483 <JIA>
A;Cross-references: UNIPROT:P33322; UNIPARC:UPI0000052ECF; EMBL:L12351; NID:g311096; PID
A;Accession: S45443
A;Molecule type: protein
A;Residues: 36-45;142-149;168-175;385-401 <JI2>
A;Cross-references: UNIPARC:UPI000017B272; UNIPARC:UPI000017B273; UNIPARC:UPI000017B274;
R;Wohldmann, P.
submitted to the EMBL Data Library, November 1994
A;Description: The sequence of S. cerevisiae cosmid 9470.
A;Reference number: S51414
A;Accession: S51420
A;Molecule type: DNA
A;Residues: 1-483 <WOH>
A;Cross-references: UNIPARC:UPI0000052ECF; EMBL:U17246; NID:g577192; PIDN:AAB67463.1; PI
C;Genetics:
A;Gene: SGD:CBF5; MIPS:YLR175w
A;Cross-references: SGD:S0004165; MIPS:YLR175w
A;Map position: 12R
C;Function:
A;Description: binds centromeres and microtubules [validated, MUID:93330283]
F;434-463/Region: Lys-Lys-Glu/Asp repeats

  Query Match             13.5%;  Score 75;  DB 2;  Length 483;
  Best Local Similarity   29.5%;  Pred. No. 23;
  Matches   33;  Conservative   15;  Mismatches   42;  Indels   22;  Gaps    6;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIP-----HQSSLPQDNRGN 57
              |  || :  |  | :  | |||:     |  |  :  |:|       :|  |: :
Db        356 AQKKKQM---KADGKL---DKYGRVN----ENTPEQWKKEYVPLDNAEQSTSSSQETKET 405

Qy         58 SRD---SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
                :   :|| |:|    :|  |   | : |  |   :||||   | |:|| |
Db        406 EEEPKKAKEDSLI----KEVETEKEEVKEDDSKKEKKEKKDKKEKKEKKEKK 453


RESULT 28
S57196
calpain (EC 3.4.22.17) large chain 3 - chicken
N;Alternate names: calpain p94 heavy chain; n-calpain-1 heavy chain
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S57196
R;Sorimachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
Biochim. Biophys. Acta 1261, 381-393, 1995
A;Title: Identification of a third ubiquitous calpain species - chicken muscle expresses
A;Reference number: S57194; MUID:95260862; PMID:7742367
A;Accession: S57196
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-810 <SOR>
A;Cross-references: UNIPROT:Q92177; UNIPARC:UPI0000126E9C; EMBL:D38028; NID:g882072; PID
C;Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C;Keywords: cysteine proteinase; EF hand; hydrolase
F;93-393/Domain: calpain catalytic domain homology <CALP>
F;746-778/Domain: calmodulin repeat homology <EFH>
```

```
  Query Match             13.5%;  Score 75;  DB 1;  Length 810;
  Best Local Similarity   21.6%;  Pred. No. 42;
  Matches   27;  Conservative   21;  Mismatches   45;  Indels   32;  Gaps    5;

Qy          8 NLERKKEHGDILAEDLYGRLEIPAIE---LPSENERGYYIPHQ---------------- 47
              |  : :    |    ::  |  :|  |   :||      | |||
Db        521 NASKARSKTYINMREISERFRLPPSEYVIIPST-----YEPHQEGEFILRVFSEKRSLSE 575

Qy         48 ---SSLPQDNRGNSRDSKEISIIE---KTNRESITTNVEGRRDIHKGHLEEKKDGSIKPE 101
                : :  |      :  | |  :     :|:| :||: :  :|  | |::|||  | |
Db        576 EVENMIEADRPSKKKKGKPIIFVSDRANSNKE-LTTDEDAGKDGEKTHVDEKKRSSAKAR 634

Qy        102 QKEDK 106
                :| ::
Db        635 EKSEE 639


RESULT 29
E71616
hypothetical protein PFB0365w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: E71616
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71616
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1247 <GAR>
A;Cross-references: UNIPROT:Q8WSS7; UNIPARC:UPI000017B605; GB:AE001390; GB:AE001362; NID
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0365w

  Query Match             13.5%;  Score 75;  DB 2;  Length 1247;
  Best Local Similarity   28.2%;  Pred. No. 69;
  Matches   31;  Conservative   15;  Mismatches   46;  Indels   18;  Gaps    5;

Qy          8 NLERKKEHGDILAEDLYGRLEIPAIELPSENERG-YYIPHQSSLPQDNRGNSRDS---KE 63
              | :| ||:      ::     |  | :  ||| |   :  |     |:|  : ||:
Db         83 NNKRSKENSGFYNSEINA---INDIMIGYENEGGKKKVQRQKKYHQEN--DERDAYKRNN 137

Qy         64 ISIIEKTNRESITTNVEGRRDIH---------KGHLEEKKDGSIKPEQKE 104
                |  ||| :  | || ::             ||  |:| :|  | ||:|
Db        138 AKYEEDDNREQVEKNNEGEKNDEGGKKGEGEKKGEGEKKGEGEKKDEQEE 187


RESULT 30
T26287
hypothetical protein W08E3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T26287
R;Dobson, R.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20189
A;Accession: T26287
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-272 <WIL>
A;Cross-references: UNIPROT:P91916; UNIPARC:UPI000017BC30; EMBL:Z92773; PIDN:CAB07130.1;
A;Experimental source: clone W08E3
C;Genetics:
A;Gene: CESP:W08E3.2
A;Map position: 1
A;Introns: 30/3; 65/1; 120/2

  Query Match             13.4%;  Score 74.5;  DB 2;  Length 272;
  Best Local Similarity   21.5%;  Pred. No. 13;
  Matches   28;  Conservative   25;  Mismatches   46;  Indels   31;  Gaps    4;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQ----------SSL 50
              :| ::||:: :   |   :    : : |  :     : |: | |:           :
Db        120 SDDRRNLKKSRADCDWKHDRFDEKFQRPKTKKQIVTKYGFDIRHKDGTPIQDGEEIDAPA 179

Qy         51 PQDNRGNSRDSKEISIIEKTNRESITTNVE------GRR----------DIHKGHLEEK 93
              |:     | |::::|  :|  | | ||  :       ||          |  ||  ||:
Db        180 PESTEEESSDARQVSRQQKDRRVSSTTRGKVGKSGTSRRPMKQREREDTVDESKG--EEE 237

Qy         94 KDGSIKPEQK 103
              :|  ::| ||
Db        238 EDQGVRPAQK 247


Search completed: December 15, 2005, 21:49:27
Job time : 42 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 20, 2005, 01:36:14 ; Search time 3240 Seconds
(without alignments)
1545.129 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09837344/runat_15122005_103706_204/app_query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_5315_@runat_15122005_103706_204 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	396	71.5	317	10	T09652	T09652 0224m3 gmbP
2	93	16.8	691	2	BG832038	BG832038 602765038
3	91.5	16.5	386	9	AZ164328	AZ164328 SP_0075_B
4	90	16.2	1042	3	BI559491	BI559491 603252811
5	90	16.2	1218	10	CL080889	CL080889 CH216-160
c 6	89.5	16.2	598	10	CE688248	CE688248 tigr-gss-
7	89	16.1	1714	10	AG390929	AG390929 Mus muscu
c 8	88.5	16.0	577	9	BZ425793	BZ425793 BONAI70TF
c 9	88.5	16.0	817	9	BH446456	BH446456 BOHKD72TF
10	88	15.9	1082	2	BG497835	BG497835 602543125
c 11	86.5	15.6	2131	10	AG332417	AG332417 Mus muscu
c 12	86	15.5	767	8	CX805523	CX805523 JGI_CAAJ1
c 13	86	15.5	800	8	DN635922	DN635922 ACAC-aab7
14	86	15.5	872	8	CX805524	CX805524 JGI_CAAJ1
15	86	15.5	959	2	BI114982	BI114982 602862806
16	85.5	15.4	1324	2	BE732599	BE732599 601571156
c 17	85	15.3	641	11	CR483922	CR483922 mth2-151G
18	85	15.3	824	2	BE643089	BE643089 Cri2_7_N0
c 19	85	15.3	890	11	CNS07CXI	AL439756 T3 end of
20	84.5	15.3	642	8	DR437990	DR437990 ACAC-aaa7
21	84.5	15.3	693	9	BH381738	BH381738 AG-ND-178
c 22	84.5	15.3	805	6	CA775281	CA775281 io87b08.x
c 23	84.5	15.3	872	10	AG872408	AG872408 Oryza sat
24	84.5	15.3	1513	10	AG453925	AG453925 Mus muscu
c 25	84	15.2	606	11	DE006900	DE006900 Branchios
26	84	15.2	1439	8	DN674862	DN674862 CFW83-A07
27	83.5	15.1	602	5	BX320026	BX320026 BX320026
c 28	83.5	15.1	621	10	CW816235	CW816235 pinea_259
c 29	83.5	15.1	633	5	BW317679	BW317679 BW317679
c 30	83.5	15.1	684	7	CO131269	CO131269 GR__Eb43H
31	83.5	15.1	770	9	BH391018	BH391018 AG-ND-155
32	83.5	15.1	816	7	CO113362	CO113362 GR__Eb013
33	83.5	15.1	1666	2	BF343790	BF343790 602015455
c 34	83	15.0	227	11	DE035928	DE035928 Branchios
35	83	15.0	295	7	CV004955	CV004955 csa01-2ms
36	83	15.0	465	9	AQ929695	AQ929695 RPCI-23-2
37	83	15.0	505	7	CN768666	CN768666 taf69g09.
c 38	83	15.0	538	7	CN622446	CN622446 tad91c02.
39	83	15.0	591	7	CN555488	CN555488 tae13d07.
c 40	83	15.0	705	7	CN555161	CN555161 tae13d07.
41	83	15.0	714	10	BX155902	BX155902 Danio rer
42	83	15.0	1045	3	BI752474	BI752474 603027754
43	82.5	14.9	360	1	AW435888	AW435888 74941 MAR
44	82.5	14.9	589	3	BM169017	BM169017 EST571540
45	82.5	14.9	605	6	CA082727	CA082727 SCBFAM202

ALIGNMENTS

RESULT 1
T09652
LOCUS T09652 317 bp DNA linear GSS 05-DEC-2000
DEFINITION 0224m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic clone 0224m, genomic survey sequence.
ACCESSION T09652
VERSION T09652.1 GI:319484
KEYWORDS GSS.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 317)
AUTHORS Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C. and Dame,J.B.
TITLE Gene sequence tags from Plasmodium falciparum genomic DNA fragments prepared by the genease activity of mung bean nuclease
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
PUBMED 8234327
COMMENT Other_GSSs: 0224m7
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .317
/organism="Plasmodium falciparum"

/mol_type="genomic DNA"
/db_xref="taxon:5833"
/clone="0224m"
/lab_host="E. coli XL1-Blue"
/clone_lib="gmbPfHB3.1, G. Roman Reddy"
/note="Vector: pBlueScript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were polished using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBlueScript SK(+). Recombinant plasmids transformed E. coli XL1-Blue."

ORIGIN

Alignment Scores:

Pred. No.:	9.02e-38	Length:	317
Score:	396.00	Matches:	77
Percent Similarity:	96.25%	Conservative:	0
Best Local Similarity:	96.25%	Mismatches:	3
Query Match:	71.48%	Indels:	0
DB:	10	Gaps:	0

US-09-837-344-19 (1-107) x T09652 (1-317)

```
Qy         28 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 47
              |||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCGAAGCTATAGAACTTCCATCAGAGAATGAACGTGGATATTATATACCACATCAA 60

Qy         48 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 67
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 120

Qy         68 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 87
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 180

Qy         88 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 107
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        181 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGTACAAAAAGAAGATAAATCT 240
```

RESULT 2
BG832038

```
LOCUS       BG832038                691 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  602765038F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4907265 5',
            mRNA sequence.
ACCESSION   BG832038
VERSION     BG832038.1  GI:14179625
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1811  row: k  column: 10
            High quality sequence stop: 45.
FEATURES             Location/Qualifiers
     source          1. .691
```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4907265"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_42"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. |"

ORIGIN

Alignment Scores:

Pred. No.:	0.946	Length:	691
Score:	93.00	Matches:	38
Percent Similarity:	44.78%	Conservative:	22
Best Local Similarity:	28.36%	Mismatches:	41
Query Match:	16.79%	Indels:	33
DB:	2	Gaps:	6

US-09-837-344-19 (1-107) x BG832038 (1-691)

```
Qy          1 ArgLysAlaAspThrLysLysAsn------LeuGluArgLysLysGlu------------ 14
              ||||||   ||||||||||||           :::||||||||||||
Db        242 CGAAAGGAAGATACAAAAAAAACACAATACAGTGATAGAAAGAAAGAAAAAAACAAAGCT 301

Qy         15 ---HisGlyAspIleLeuAlaGluAsp-----------------LeuTyrGlyArgLeu 27
                 |||     |||:::    |||                      ::::::
Db        302 ATACATAGCAAAATCATAGACAGAGATCAAAACGAGACGAACAACAAACAAACACAAGTA 361

Qy         28 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 47
              :::|||      ||||||   |||   :::|||::::::
Db        362 GACATACAAACAATAGAACAACCGAAGAAAAATAAGAAA--------------ATAAAC 406

Qy         48 SerSerLeuProGlnAspAsnArgGlyAsnSer-------------------------- 58
                 :::|||   :::|||   :::   |||:::
Db        407 AAGACACTAGAAGAGGACAAAAAAACAAACAACACACGAAAACATACAAAACACCACAAG 466

Qy         59 ------ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGlu--------- 73
                  :::     ||||||   :::      :::|||     |||:::
Db        467 AACGAACAGACAAGAAAAGAAAGAACAAAAACAAAAAAGAATCAAAGAGACAGAACGTAG 526

Qy         74 SerIleThrThrAsnValGluGly-ArgArgAspIleHisLysGlyHisLeuGluGluLy 93
              |||   |||   |||   |||    |||::::::|||   |||          |||:::
Db        527 AGCAAAACAAGAAACAAAGAAGATAAGAAAAGAGATACAAAAGAAAGAACAAGAACAAAA 586

Qy         93 sLysAspGlySerIleLysProGluGlnLysGluAspLys 106
              |||:::         |||   |||:::||||||   :::
Db        587 TAAAGAAACAAAACATAAAGACGAAAAAAAAGAAACAAGA 626
```

RESULT 3
AZ164328

```
LOCUS       AZ164328                386 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  SP_0075_B1_G05_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=75 Col=9 Row=N, genomic survey sequence.
ACCESSION   AZ164328
VERSION     AZ164328.1  GI:8317045
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
```

```
            Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H.
            and Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
   PUBMED   10920195
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 75  row: N  column: 9
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 386.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7668"
                     /clone="Plate=75 Col=9 Row=N"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
ORIGIN

Alignment Scores:
Pred. No.:                0.695          Length:         386
Score:                    91.50          Matches:        20
Percent Similarity:       59.65%         Conservative:   14
Best Local Similarity:    35.09%         Mismatches:     18
Query Match:              16.52%         Indels:         5
DB:                       9              Gaps:           2

US-09-837-344-19 (1-107) x AZ164328 (1-386)

Qy          35 ProSerGluAsnGluArgGlyTyr---TyrIleProHisGlnSerSerLeu--------- 50
               ||||||          |||   |||   ::::::|||       |||||||
Db          68 CCGTCTGCACTTAGGCGATTCTATGATTTTGTTCCGCGGAATTCATCTTTAAGATGAGAG 127

Qy          51 ---ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLys 69
                  :::   |||:::|||      :::      ||||||     |||::::::|||
Db         128 TCCACAAAAAAAAACAAAGGAAAAAAAAAATTTAGGAAAGAAAGATGTATATTGAAAAAG 187

Qy          70 ThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis 86
                  |||   :::|||:::      |||::::::|||:::::::::|||||
Db         188 GTGAACTTGAAAAGCTTAAAATGCAATATTGATGGCAAAAAGGAAATACAT 238

RESULT 4
BI559491
LOCUS       BI559491                1042 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603252811F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5295338 5',
            mRNA sequence.
ACCESSION   BI559491
VERSION     BI559491.1  GI:15446805
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1042)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11747  row: e  column: 03
            High quality sequence stop: 12.
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5295338"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_97"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.2 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                3.64           Length:         1042
Score:                    90.00          Matches:        36
Percent Similarity:       40.16%         Conservative:   15
Best Local Similarity:    28.35%         Mismatches:     54
Query Match:              16.25%         Indels:         22
DB:                       3              Gaps:           5

US-09-837-344-19 (1-107) x BI559491 (1-1042)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAsp---IleLeu 19
               |||:::      |||||||||      ||||||::::::             :::
Db         498 CGACGAAAAACGACCAAAAAGCAAACAGAGAGACAAAGGAGCAGAAGGCAGAGAGTAAAA 557

Qy          20 AlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGlu 39
               ||||||     ||||||                      |||   |||::::::
Db         558 GAGGAAGACGAGCGAGGAAGGAGATGAAGAACGGGGAGGGAA---CCAGCACAAAGACGA 614

Qy          40 ArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArg------GlyAsn 57
               |||          :::|||        ||||||   :::              
Db         615 GAGGGGAAGAGCAAAACAAGAGAAAGCAAAGAGAGACAGGACAAGAAAACACAGGAAGAA 674

Qy          58 SerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThr 77
               |||:::   |||:::            ||||||  :::|||         ||||||
Db         675 GAAAGAGAGAAAAAACAAACACAAAAAGACGAAAAGAGACACAGAACACCGAGCACGACA 734

Qy          78 AsnValGluGlyArgArgAspIleHisLys---------------------------- 87
                  :::|||:::|||      ||||||
Db         735 GAAGAAAAGGGAAAACGAAGAAGACACAAGACACCAGGGAAACGGAGAGCACAGAGGAAA 794

Qy          88 ---GlyHis--------------------LeuGluGluLysLysAspGlySerIleLys 99
               ||||||                    ||||||:::||||||
Db         795 GACGGACACGGAGGAGAAGAAGGGAGAAAGAAAGAAGAAAGAAAAGACAAAGAGAAACAC 854

Qy         100 ProGluGlnLysGluAspLys 106
               |||:::   |||   |||
Db         855 AAAGAAGAACACGAACAAAAG 875

RESULT 5
CL080889
LOCUS       CL080889                1218 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-160D11_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-160D11, genomic survey sequence.
ACCESSION   CL080889
VERSION     CL080889.1  GI:40536802
KEYWORDS    GSS.
```

```
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1218)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 82
            High quality sequence stop: 134.
FEATURES             Location/Qualifiers
     source          1. .1218
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-160D11"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

Alignment Scores:
Pred. No.:              4.42          Length:        1218
Score:                  90.00         Matches:       30
Percent Similarity:     51.95%        Conservative:  10
Best Local Similarity:  38.96%        Mismatches:    30
Query Match:            16.25%        Indels:        7
DB:                     10            Gaps:          4

US-09-837-344-19 (1-107) x CL080889 (1-1218)

Qy          30 ProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHis---GlnSer 48
               ||||||          |||   :::|||   :::      :::   ||||||   |||:::
Db         434 CCAGCAGCACCCCCACCACAAAAAAACAACCAGANCNGNCACAGCCCGCACACACAGACC 493

Qy          49 SerLeuProGlnAspAsnArgGlyAsn-SerArgAspSerLysGluIleSerIleIleGl 68
               |||||||||       |||    |||              :::             ||
Db         494 CAGCAGAGACAAGACAACACAACCAACCGGCAGAAGCAAA---------GCGACCCAAGA 544

Qy          68 uLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGl 88
               |:::      |||||||||       |||      ||||||||:::      |||||||||
Db         545 GGAGCAACAGAGGGAAAGC---AAGACACGAAGAGAAGGACGCCAGAAAACGCACAAAGG 601

Qy          88 yHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGlu 104
               |       :::||||||       |||      ||||||:::|||:::|||
Db         602 AAAA---AAGGAAAAACCCAGGGGAGGAGAAAAGCCAAAACAGCAGGAA 647

RESULT 6
CE688248/c
LOCUS       CE688248                598 bp    DNA     linear   GSS 29-SEP-2003
DEFINITION  tigr-gss-dog-17000368528982 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE688248
VERSION     CE688248.1  GI:37007318
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  PUBMED    14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:              2.09          Length:        598
Score:                  89.50         Matches:       34
Percent Similarity:     48.96%        Conservative:  13
Best Local Similarity:  35.42%        Mismatches:    34
Query Match:            16.16%        Indels:        15
DB:                     10            Gaps:          5

US-09-837-344-19 (1-107) x CE688248 (1-598)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
               ||||||   :::   ||||||       ||||||||||||         :::|||||||||
Db         490 AGAAAGAAAGAAAGAAAGAAA------GAAAGAAAGAAAAGAAAAAAAGAAATCCTGGCA 437

Qy          21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeu-----------ProSer 36
                  :::         :::|||:::      |||  |||                  |||
Db         436 TTTGCAATGACATGG---ATGGAACTAGAGGATATTATGCTAAGTAAAATAAGTAAATCA 380

Qy          37 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu---------ProGlnAsp 53
               |||||||||       |||   :::   |||          |||             :::|||
Db         379 GAGAATGAATTATCATATGATCTCACTCACATGTGGAAACTAAGAAACAAAACAGAGGAT 320

Qy          54 AsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGlu 73
               :::||||||       |||:::      |||   ::::::||||||   :::
Db         319 CATAGGGGAAGAATAGGAAAAATAAAACAAGATGAAATCAGAAAGGAGACAAACCATAAG 260

Qy          74 SerIle--------------ThrThrAsnValGluGlyArgArgAsp 84
               |||:::                  |||   ||||||||||||
Db         259 AGCCTCATAATCATAGGAAACAAACTGAAGGTTGCTGGAAGGAGGGAT 212

RESULT 7
AG390929
LOCUS       AG390929                1714 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-207N12.TJ, genomic survey
            sequence.
ACCESSION   AG390929
VERSION     AG390929.1  GI:48009507
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
```

genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1714)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institude, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@rtc.riken.jp PRIMERS Sequencing : TJ LIBRARY Vector : pBACe3.6 R.Site 1 : EcoRI R.Site 2 : EcoRI.

FEATURES Location/Qualifiers
source 1. .1714
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-207N12.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Alignment Scores:

Pred. No.:	8.93	Length:	1714
Score:	89.00	Matches:	29
Percent Similarity:	43.86%	Conservative:	21
Best Local Similarity:	25.44%	Mismatches:	46
Query Match:	16.06%	Indels:	18
DB:	10	Gaps:	3

US-09-837-344-19 (1-107) x AG390929 (1-1714)

```
Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||   :::   |||:::      |||   ||||||||||:::|||
Db       1282 AGGAAAGAGGAGAAGAAGGAGAGAAGAGAAGAAAAGAAAGAAAATGGG------AGAGAA 1335

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
              |||         ||||||   |||:::         |||         |||   ::::::
Db       1336 GAGAAGAGAGGAAGAAGAAGAAGTACAAGAAAGAGAAAGAAGAGAGGAAAAAAAGAAG 1395

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
              |||                          ::::::   ||||||      |||
Db       1396 GGAAGAAAGAAAGGAGAGGGAAAGGATAAGAAGAAAGAAGGAAGGGGAAAGAAAAGA--- 1452

Qy         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80
                                  :::   |||::::::   :::             |||
Db       1453 --------------------AGAAGGAGAAATAAGAAAGAGATGAGAGAGAGAAGAGAA 1491

Qy         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer--------- 97
                :::|||:::      :::|||   :::||||||:::|||
Db       1492 GAAAAAAGAGAGAAAGGGAGAGGAAAAATAGAAGAAGAGAAAGGGGAAAAGAGGAGAAAA 1551

Qy         98 --------------IleLysProGluGlnLysGluAspLys 106
                           ||||||   |||::::::|||::::::
Db       1552 GAAAGAAAGTGAAGAATAAAGAGGGAGAGAAGGGAAGAAAGA 1593
```

RESULT 8
BZ425793/c

LOCUS BZ425793 577 bp DNA linear GSS 13-DEC-2002
DEFINITION BONAI70TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONAI70, genomic survey sequence.
ACCESSION BZ425793
VERSION BZ425793.1 GI:26667305
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 577)
AUTHORS Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
TITLE Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis
JOURNAL Genome Res. 15 (4), 487-495 (2005)
PUBMED 15805490
COMMENT Other_GSSs: BONAI70TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES Location/Qualifiers
source 1. .577
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONAI70"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Alignment Scores:

Pred. No.:	2.64	Length:	577
Score:	88.50	Matches:	20
Percent Similarity:	51.02%	Conservative:	5
Best Local Similarity:	40.82%	Mismatches:	21
Query Match:	15.97%	Indels:	3
DB:	9	Gaps:	1

US-09-837-344-19 (1-107) x BZ425793 (1-577)

```
Qy         11 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 30
              |||   :::   ||||||      ||||||            ||||||   :::   |||
Db       231 CGTCACCAGTCCCATGGGCCTGCCCTTGCG---------CTTGGCCGTAGTCAGGAACCT 181

Qy         31 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 50
                 :::|||||||||||||||:::      |||      ||||||
Db       180 TGCTCACAACTTCCCTCAGAAAATCAAACACGTTATATCTTTCCACACAGTGTTTTCTGT 121

Qy         51 ProGlnAspAsnArgGlyAsnSerArg 59
                 |||   |||:::   |||   |||
Db       120 AAACAACAAAACAAGTTCAATAAGCGT 94
```

RESULT 9
BH446456/c
LOCUS BH446456 817 bp DNA linear GSS 12-DEC-2001
DEFINITION BOHKD72TF BOHK Brassica oleracea genomic clone BOHKD72, genomic survey sequence.
ACCESSION BH446456
VERSION BH446456.1 GI:17632170
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
            Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea and its
            application to gene discovery and annotation in Arabidopsis
  JOURNAL   Genome Res. 15 (4), 487-495 (2005)
   PUBMED   15805490
COMMENT     Other_GSSs: BOHKD72TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHKD72"
                     /clone_lib="BOHK"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Alignment Scores:
Pred. No.:              4.07          Length:         817
Score:                  88.50         Matches:        20
Percent Similarity:     51.02%        Conservative:   5
Best Local Similarity:  40.82%        Mismatches:     21
Query Match:            15.97%        Indels:         3
DB:                     9             Gaps:           1

US-09-837-344-19 (1-107) x BH446456 (1-817)

Qy          11 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 30
               |||   :::   ||||||      ||||||            ||||||   :::   |||
Db         596 CGTCACCAGTCCCATGGGCCTGCCCTTGCG---------CTTGGCCGTAGTCAGGAACCT 546

Qy          31 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 50
               :::||||||||||||||:::      |||      ||||||
Db         545 TGCTCACAACTTCCCTCAGAAAATCAAACACGTTATATCTTTCCACACAGTGTTTTCTGT 486

Qy          51 ProGlnAspAsnArgGlyAsnSerArg 59
               |||   |||:::   |||   |||
Db         485 AAACAACAAAACAAGTTCAATAAGCGT 459

RESULT 10
BG497835
LOCUS       BG497835               1082 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  602543125F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4665487 5',
            mRNA sequence.
ACCESSION   BG497835
VERSION     BG497835.1  GI:13459352
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1082)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1471  row: a  column: 08
            High quality sequence stop: 173.
FEATURES             Location/Qualifiers
     source          1. .1082
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4665487"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_60"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); Double-stranded cDNA was prepared from
                     cell line RNA.  5' and 3' adaptors were used in cloning as
                     follows: 5' adaptor sequence:  5'-CACGGCCATTATGGCC-3' and
                     3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.5
                     kb (range 0.9-4.0 kb). 14/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              6.65          Length:         1082
Score:                  88.00         Matches:        32
Percent Similarity:     41.96%        Conservative:   15
Best Local Similarity:  28.57%        Mismatches:     33
Query Match:            15.88%        Indels:         32
DB:                     2             Gaps:           4

US-09-837-344-19 (1-107) x BG497835 (1-1082)

Qy           1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
               ::::::    |||||||||      |||   :::|||:::   |||:::      |||
Db         606 AAGCGGAGGCGAACGAAGAAGAGAAGGGAAGGAGAAAAAGACGGGGGGGAGAAGAAGGCA 665

Qy          21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
               :::                                                    ||||||
Db         666 AAG-------------------------------------------------GAAAGA 674

Qy          41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn------ArgGlyAsnSer 58
               |||                  |||         ::::::|||      ||||||   |||
Db         675 GGAGAGGAGGCAGAAGAAGGAAGCGAGCAGGAAAGAGAGAACACAGAGAGAGGAGAGAGT 734

Qy          59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSer------------ 74
               |||                  ||||||     |||||||||
Db         735 AGAAGA--------------------GAGAAAAAAAGAAGAGAGAGCGAAGAGAGGAAA 773

Qy          75 ------IleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGlu 92
                     |||      :::|||||||||          ||||||      :::|||
Db         774 GGAAAAGCGGAGACAGAGGAAAAAGGAAGAAGAAGCAGGAAGAAGGGAAAGAGGAAAGAA 833

Qy          93 LysLysAspGlySerIleLysProGluGlnLysGlu 104
               :::|||:::|||      :::   |||::::::|||
Db         834 AGGAAAGAAGGGAAGAGGAGGAAGGAGAAGAGAGAG 869

RESULT 11
AG332417/c
LOCUS       AG332417               2131 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-122J15.T7, genomic survey
            sequence.
ACCESSION   AG332417
VERSION     AG332417.1  GI:47905727
KEYWORDS    GSS.
```

```
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2131)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .2131
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-122J15.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:              23.5          Length:        2131
Score:                  86.50         Matches:       42
Percent Similarity:     44.35%        Conservative:  13
Best Local Similarity:  33.87%        Mismatches:    38
Query Match:            15.61%        Indels:        32
DB:                     10            Gaps:          7

US-09-837-344-19 (1-107) x AG332417 (1-2131)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGlu------------------ 14
              ||||||   :::   :::|||      ||||||:::||||||
Db       1165 AGAAAAGAGAACCCACAGAAGAGAAACGAAAGAGAAAAAGAAAGAGATCAAGAAAGACGA 1106

Qy         15 ------------------HisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGlu 28
                               |||       |||:::   :::       |||   |||
Db       1105 GAAACGATAAATAAACGACAAGGCACAACGAGAGCACAAAAGATGGGAGACAGAGCAGAG 1046

Qy         29 IleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSer 48
              ||||||:::   |||         |||      |||
Db       1045 ATACCGAGCAA-GAGAGAAGAGACGAGAGGAGAAGA---------------------- 1011

Qy         49 SerLeuProGlnAspAsnArgGlyAsnSerArg---------AspSerLysGluIleSer 65
                          |||   ||||||:::   |||             ||||||
Db       1010 ------CAAAGCGACGAGAGAGGAAGCGGACGAAGAAAAAATACAAGGAAGGAAGACGAT 957

Qy         66 IleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIle 85
              ||||||||||||:::|||      ||||||:::   :::   |||:::|||
Db       956 GACAAAGAGAAGACGAACCAAGAA------ACGACACACCATAAGAACAGGAAAGACAGA 903

Qy         86 HisLysGlyHisLeu------GluGluLysLysAspGly---SerIleLysProGluGln 102
                 |||   |||         ||||||||||||||      |||   |||      |||
Db       902 CAAAAA---CACACAAAAAGAGAAGAGAAAAAGGACAGCATAAGCACAAAAAGGAGACAA 846

Qy        103 LysGluAspLys 106
                 |||   ::::::
Db       845 AAAAACAACGAG 834

RESULT 12
CX805523/c
LOCUS       CX805523                767 bp    mRNA    linear   EST 28-JUL-2005
DEFINITION  JGI_CAAJ17255.rev NIH_XGC_tropBrn2 Xenopus tropicalis cDNA clone
            IMAGE:7647188 3', mRNA sequence.
ACCESSION   CX805523
VERSION     CX805523.2  GI:71461842
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C.,
            Brokstein,P. and Lindquist,E.A.
  TITLE     DOE Joint Genome Institute Xenopus tropicalis EST project
  JOURNAL   Unpublished (2004)
COMMENT     On Jan 31, 2005 this sequence version replaced gi:58360150.
            Other_ESTs: JGI_CAAJ17255.fwd
            Contact: Lindquist,E.A., Richardson,P.
            DOE Joint Genome Institute
            2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            Tel: 925 296 5600
            Fax: 925 296 5710
            Email: cdna@jgi-psf.org
            Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,
            University of California, Berkeley:
            http://tropicalis.berkeley.edu/home)
            cDNA Library Preparation: DOE Joint Genome Institute:
            http://www.jgi.doe.gov
            DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
            Clone Distribution: I.M.A.G.E. Consortium/LLNL:
            http://image.llnl.gov
            Naming Conventions: EST name is generated by the concatenation of
            the JGI Clone Id and the direction of sequencing. The suffix '.rev'
            indicates a reverse sequencing read of the insert. It does not
            necessarily reflect the orientation of the insert.
            Poly-A: Based upon the presence of a run of 14 or more T residues
            at the beginning of the sequence, this clone was polyadenylated.
            The resulting Poly-T sequence has been removed.
            Plate: CAAJ 0177  row: n  column: 18
            High quality sequence stop: 764
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:7647188"
                     /tissue_type="Brain"
                     /dev_stage="Adult"
                     /lab_host="ElectroMAX DH10B"
                     /clone_lib="NIH_XGC_tropBrn2"
                     /note="Vector: pCMVsport6; Site_1: SalI; Site_2: NotI;
                     This library was made from dT primed cDNA and cloned into
                     Invitrogen pCMVsport6 vector. The work was done at DOE
                     Joint Genome Institute. Poly A RNA were primed with 5'
                     GACTAGTTCTAGATCGCGAG CGGCCGCCCTTTTTTTTTTTTTT 3'. CDNA
                     were ligated to SalI adapter (5' TCGACCCACGCGTCCG and
                     5'CGGACGCGTGGG), digested with NotI, size fractionated in
```

1.1% agarose gel electrophoresis and ligated into NotI and SalI digested pCMVsport6 vector."
ORIGIN

Alignment Scores:

Pred. No.:	7.55	Length:	767
Score:	86.00	Matches:	18
Percent Similarity:	54.84%	Conservative:	16
Best Local Similarity:	29.03%	Mismatches:	18
Query Match:	15.52%	Indels:	10
DB:	8	Gaps:	1

US-09-837-344-19 (1-107) x CX805523 (1-767)

```
Qy         32 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 51
              |||   :::|||   ::::::   |||      |||:::||||||   ||||||||||||
Db        761 ATTTCCATTCCAGTGGACAGCCCTCGGAGCACTTACTTACCTCATGCTTCTAGCCTTCCT 702

Qy         52 GlnAspAsnArgGlyAsnSerArgAspSerLysGlu---------------------- 63
              :::|||::::::           :::||||||
Db        701 GAAGATAGCAAATCAAAGTACAATGAATCAAAAACAAACGTACGGAGAATATATGTTGAG 642

Qy         64 ------IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGly 81
                    ::::::   :::   ::::::::::::   |||   ||||||:::|||
Db        641 CCTTACAAACAAATGATGTCTGAGGAAGACCAGAAAACGTGGACCTTAAATGTAGATGGA 582

Qy         82 ArgArg 83
                |||
Db        581 TGCCGT 576
```

RESULT 13
DN635922/c

```
LOCUS       DN635922                800 bp    mRNA    linear   EST 30-JUN-2005
DEFINITION  ACAC-aab75d19.b1 Hydra EST UCI 7 Hydra magnipapillata cDNA 3', mRNA
            sequence.
ACCESSION   DN635922
VERSION     DN635922.2  GI:68387167
KEYWORDS    EST.
SOURCE      Hydra magnipapillata
  ORGANISM  Hydra magnipapillata
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
            Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,
            Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
            Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R.,
            Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
  TITLE     WashU Hydra EST Project
  JOURNAL   Unpublished (2002)
COMMENT     On Mar 28, 2005 this sequence version replaced gi:61936718.
            Contact: Hans Bode
            WashU Hydra EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library material provided by Hans Bode & Dirk Lindgens, Univ. of
            Calif., Irvine Library constructed by Dirk Lindgens, Univ. of
            Calif. Irvine Library sequenced by Washington University Genome
            Sequencing Center For information on obtaining a clone please
            contact: Hans Bode (hrbode@uci.edu)
            This trace has been recalled with phred
            original value before phred recall for SL was 9
            original value before phred recall for SR was 921
            Seq primer: degenerate primer.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Hydra magnipapillata"
                     /mol_type="mRNA"
                     /strain="105"
                     /db_xref="taxon:6085"
                     /lab_host="DH10B"
                     /clone_lib="Hydra EST UCI 7"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; a.
                     1st strand cDNA was primed with a Not I primer-adapter
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3') b.
                     Double-stranded cDNA was ligated to Sal I adapter,
                     digested with Not I and cloned into the pSPORT 1-vector
                     pre-cut with Not I and Sal I. c. The ligation mix was
                     transformed into DH10B cells. d. The cells were grown in
                     SOC = 5g yeast, 20g tryptone, 0.5 g NaCl, 10 mM MgSO4, 10
                     mM MgCl, 0.2% glucose / Liter, (no antibiotic). e. DMSO
                     was added to a final conc. of 10% as a cryoprotectant.and
                     frozen. f. The titre before freezing was determined as ~
                     6000/ 100 ul. Assuming a 10% loss upon freezing, the titre
                     is probably ~ 5400/ 100 ul. g. 6 tubes each containing ~
                     5400 clones/ul [ = total of ~ 32,400] are enclosed. h.
                     The frequency of vectors containing inserts is 95% as
                     determined by digestion check after picking 20 clones,
                     miniprep and subsequent digestion with Not I and Sal I. i.
                     A low level of 32P was used in the cDNA synthesis
                     procedure. The level measured by holding a Geiger Counter
                     next to a tube was background."
ORIGIN
```

Alignment Scores:

Pred. No.:	7.95	Length:	800
Score:	86.00	Matches:	31
Percent Similarity:	42.40%	Conservative:	22
Best Local Similarity:	24.80%	Mismatches:	46
Query Match:	15.52%	Indels:	26
DB:	8	Gaps:	4

US-09-837-344-19 (1-107) x DN635922 (1-800)

```
Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLysLys---------------------- 13
              |||   ||||||:::|||      :::   |||:::
Db        602 AAAGAAGACACAGAGAAAACCTCTGATGAGAAAGAACTTGCAACAGATGACTCTACACTT 543

Qy         14 --------------------GluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArg 26
                                   |||:::        :::|||:::
Db        542 ACTGATGATGATGCAAAAAGTATGAAAGGTAACCAAGATGATCAAGATGTTGATGAAGAT 483

Qy         27 LeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHis 46
              ||||||            :::      |||::::::            |||      :::
Db        482 CTTGAAGCAGAAGATGATGATTCTGAATCAAAAGACCTTGACGAAGAGTATGCTGAAAAC 423

Qy         47 GlnSerSerLeuProGlnAspAsnArg------GlyAsnSerArgAspSerLysGluIle 64
                        :::|||   :::      ||||||:::::::::::::|||::::::
Db        422 AACGACGACAACGAAGAAGACACCAAAACAACCGGGAATAACAAGGAAAATAAAGATGTC 363

Qy         65 SerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAsp 84
                           |||:::|||||||||      |||   |||
Db        362 -----------------AACCAGGAAAGTATAGATGTCAATGCAGAAAACGAAAATTCT 321

Qy         85 IleHisLysGlyHisLeuGlu---------GluLysLysAspGlySerIleLysProGlu 101
                        ::::::|||         |||||||||      ||||||      |||
Db        320 GAAGAAAATGAAAACATTGAAAATGATAATGAGAAAAAGTCTGATTCAATATCTGAAGAA 261

Qy        102 GlnLysGluAspLys 106
              |||:::||||||
Db        260 GATAAAGATGATAAA 246
```

RESULT 14
CX805524

```
LOCUS       CX805524                872 bp    mRNA    linear   EST 31-JAN-2005
DEFINITION  JGI_CAAJ17255.fwd NIH_XGC_tropBrn2 Xenopus tropicalis cDNA clone
            IMAGE:7647188 5', mRNA sequence.
ACCESSION   CX805524
VERSION     CX805524.1  GI:58360151
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C.,
            Brokstein,P. and Lindquist,E.A.
  TITLE     DOE Joint Genome Institute Xenopus tropicalis EST project
  JOURNAL   Unpublished (2004)
COMMENT     Other_ESTs: JGI_CAAJ17255.rev
            Contact: Lindquist,E.A., Richardson,P.
            DOE Joint Genome Institute
            2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            Tel: 925 296 5600
            Fax: 925 296 5710
            Email: cdna@jgi-psf.org
            Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,
            University of California, Berkeley:
            http://tropicalis.berkeley.edu/home)
            cDNA Library Preparation: DOE Joint Genome Institute:
            http://www.jgi.doe.gov
            DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
            Clone Distribution: I.M.A.G.E. Consortium/LLNL:
            http://image.llnl.gov
            Naming Conventions: EST name is generated by the concatenation of
            the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
            indicates a forward sequencing read of the insert. It does not
            necessarily reflect the orientation of the insert.
            Plate: CAAJ 0177  row: n  column: 18
            High quality sequence stop: 766.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:7647188"
                     /tissue_type="Brain"
                     /dev_stage="Adult"
                     /lab_host="ElectroMAX DH10B"
                     /clone_lib="NIH_XGC_tropBrn2"
                     /note="Vector: pCMVsport6; Site_1: SalI; Site_2: NotI;
                     This library was made from dT primed cDNA and cloned into
                     Invitrogen pCMVsport6 vector. The work was done at DOE
                     Joint Genome Institute. Poly A RNA were primed with 5'
                     GACTAGTTCTAGATCGCGAG CGGCCGCCCTTTTTTTTTTTTTTT 3'. CDNA
                     were ligated to SalI adapter (5' TCGACCCACGCGTCCG and
                     5'CGGACGCGTGGG), digested with NotI, size fractionated in
                     1.1% agarose gel electrophoresis and ligated into NotI and
                     SalI digested pCMVsport6 vector."
ORIGIN

Alignment Scores:
Pred. No.:                8.86          Length:          872
Score:                    86.00         Matches:         18
Percent Similarity:       54.84%        Conservative:    16
Best Local Similarity:    29.03%        Mismatches:      18
Query Match:              15.52%        Indels:          10
DB:                       8             Gaps:            1

US-09-837-344-19 (1-107) x CX805524 (1-872)

Qy         32 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 51
              |||   :::|||   ::::::   |||       |||:::||||||   ||||||||||||
Db        227 ATTTCCATTCCAGTGGACAGCCCTCGGAGCACTTACTTACCTCATGCTTCTAGCCTTCCT 286

Qy         52 GlnAspAsnArgGlyAsnSerArgAspSerLysGlu----------------------- 63
              :::|||::::::           :::||||||
Db        287 GAAGATAGCAAATCAAAGTACAATGAATCAAAAACAAACGTACGGAGAATATATGTTGAG 346

Qy         64 ------IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGly 81
                    ::::::   :::   ::::::::::::   |||   |||||||:::|||
Db        347 CCTTACAAACAAATGATGTCTGAGGAAGACCAGAAAACGTGGACCTTAAATGTAGATGGA 406

Qy         82 ArgArg 83
                |||
```

```
Db         407 TGCCGT 412

RESULT 15
BI114982
LOCUS       BI114982                959 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602862806F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5022073 5',
            mRNA sequence.
ACCESSION   BI114982
VERSION     BI114982.1  GI:14565883
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 959)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1841  row: k  column: 02
            High quality sequence stop: 264.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5022073"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:                9.97          Length:          959
Score:                    86.00         Matches:         33
Percent Similarity:       44.04%        Conservative:    15
Best Local Similarity:    30.28%        Mismatches:      43
Query Match:              15.52%        Indels:          18
DB:                       2             Gaps:            4

US-09-837-344-19 (1-107) x BI114982 (1-959)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              :::|||         ||||||      |||   :::   :::   |||
Db        571 AAAAAACCCCCAGAAAAGAAATAACGAGAGGTGGAACCACAAAGAGGAAAATGCCACACA 630

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40
                 |||      |||:::   |||   |||                     |||:::|||
Db        631 CGAGATACCAAAGGGAAGGGGGAACACCCA--------------------AACAAAAGG 669

Qy         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
                                        |||      ::::::   :::|||||:::|||
Db        670 ----------------------AGCGAAGGAAAGAACAAAAAGGGAAACAACAGGAGA 705

Qy         61 SerLysGluIleSerIleIleGluLysThrAsn------ArgGluSerIleThrThrAsn 78
                 |||:::         ||||||||||||:::      |||||||||
```

```
Db        706 GAAAAAAAAGCGCACCAGATAGAAAAGACACACCACCAGAGAGAATCACCACGCGAAAAA 765

Qy         79 Val---GluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySer 97
              |||   |||   ||||||     |||   |||      ::::::      ||||||
Db        766 GTGGGAGAGCAAAGACGAAAAAACCACGCGGGCGACCAACAGAAAAACAATGATGGAAGG 825

Qy         98 IleLysProGluGlnLysGluAspLys 106
                  |||:::|||      :::
Db        826 CACCACACGGAGGAAAAAAACCAAGAG 852

RESULT 16
BE732599
LOCUS       BE732599               1324 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601571156F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925861 5',
            mRNA sequence.
ACCESSION   BE732599
VERSION     BE732599.1  GI:10146591
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1324)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM752  row: o  column: 14
            High quality sequence stop: 62.
FEATURES             Location/Qualifiers
     source          1. .1324
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3925861"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:              17.1          Length:        1324
Score:                  85.50         Matches:       30
Percent Similarity:     48.31%        Conservative:  27
Best Local Similarity:  25.42%        Mismatches:    44
Query Match:            15.43%        Indels:        17
DB:                     2             Gaps:          5

US-09-837-344-19 (1-107) x BE732599 (1-1324)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeu--- 19
              :::|||         ::::::      ||||||::::::|||   |||   :::|||
Db        464 CAAAAAAGAAGAGAGAGGGAGAGAGAAGAGAGGAGGAGAGAGAGAGGCACACTCCTCTCT 523

Qy         20 ------AlaGluAspLeuTyrGlyArgLeuGluIlePro---------AlaIleGluLeu 34
                  |||:::       ||||||         |||              |||
Db        524 CTCCGCGCGCAGGGCGAGACGGGCCGCGCGTCCACACCCACACCACACACAGCAGAGAGA 583

Qy         35 ProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn 54
                 |||      :::|||   ||||||:::   |||:::           :::
Db        584 GAGAGCAGAGAGAAGAGAGACTATTATCTCTCACACAGAGAA---------GAGAGCAGA 634

Qy         55 ArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLys-ThrAsnArgGluSe 74
              :::   |||:::||||||   :::|||          |||::: :::   |||
Db        635 CAACACAACACCAGAGACGGGAGAGAG---------ACAGAGAGAGAGCGCGCGGAGAGA 685

Qy         74 rIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLy 94
                    |||:::         |||         |||:::         :::|||:::::
Db        686 CGACAACACACACACCAGCAGCAGAGAGAGAAAACATCGCGCGCAGTGCGACGAGCGACG 745

Qy         94 sAsp-Gly---------SerIleLysProGluGlnLysGluAspLysSer 107
              :::: |||         :::   :::   |||:::||||||   ::::::
Db        746 AGAGAGGGAGAGACCCCAACGCGAGAGTAGAAGAGAAAGAGCAGAGAGCG 795

RESULT 17
CR483922/c
LOCUS       CR483922                641 bp    DNA     linear   GSS 17-NOV-2004
DEFINITION  mth2-151G18FM1 BAC end, cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR483922
VERSION     CR483922.1  GI:48645498
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTH2"
                     /note="Vector: pBeloBAC11 ; Site_1: HindIII ; Site_2:
                     HindIII ; Cook, D.R. and Kim, D.J
                     Genoscope sequence ID : mth2-151G18FM1"
ORIGIN

Alignment Scores:
Pred. No.:              7.97          Length:        641
Score:                  85.00         Matches:       21
Percent Similarity:     44.87%        Conservative:  14
Best Local Similarity:  26.92%        Mismatches:    37
Query Match:            15.34%        Indels:        6
DB:                     11            Gaps:          1

US-09-837-344-19 (1-107) x CR483922 (1-641)

Qy          9 LeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGlu 28
              ||||||   |||   :::   |||   |||:::   :::         |||:::||||||
Db        426 CTAGAAACAAAAAATAAGCAGGGGCATATTATCAAAGACACCACGACAGGAAAATTAGAG 367

Qy         29 IleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSer 48
              :::         :::   |||       |||
Db        366 GTGAAGAATCCAAAACCTCCCGACCGTAATAGTTTGAACTTG----------------- 325

Qy         49 SerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGlu 68
              :::      |||:::|||     |||   |||||||||   :::::::::
Db        324 AATGAGTCTCAAGAAAACAATAATAACTACAGGGATAGTGGAGACTTGAATTATAGCAAC 265

Qy         69 LysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHis 86
              ||||||      :::   :::|||                     |||:::|||
```

```
Db        264 AAGACTCAAGTGGATGAAGTGACCATGGGGGCTACCCATGACACAGATATGCAT 211

RESULT 18
BE643089
LOCUS       BE643089                824 bp    mRNA    linear   EST 01-SEP-2000
DEFINITION  Cri2_7_N08_SP6 Ceratopteris Spore Library Ceratopteris richardii
            cDNA clone Cri2_7_N08 5', mRNA sequence.
ACCESSION   BE643089
VERSION     BE643089.1  GI:9960760
KEYWORDS    EST.
SOURCE      Ceratopteris richardii
  ORGANISM  Ceratopteris richardii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Moniliformopses; Filicophyta; Filicopsida; Filicales; Pteridaceae;
            Ceratopteris.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   Chatterjee,A., San Miguel,P., Stout,S.C., Banks,J. and Roux,S.J.
  TITLE     Expressed sequence tags of cDNA clones from a C. richardii library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Roux SJ
            Section of Molecular Cell and Developmental Biology
            University of Texas
            Biology Building, Room 16, Austin, TX 78712, USA
            Tel: 512 471 4238
            Fax: 512 232 3402
            Email: sroux@uts.cc.utexas.edu
            Plate: Cri2_7  row: N  column: 08
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Ceratopteris richardii"
                     /mol_type="mRNA"
                     /cultivar="Brogn"
                     /db_xref="taxon:49495"
                     /clone="Cri2_7_N08"
                     /tissue_type="Gametophyte"
                     /cell_type="Spore"
                     /dev_stage="20 hours after germination initiation"
                     /clone_lib="Ceratopteris Spore Library"
                     /note="Vector: pCMVSPORT6; EST sequence from cDNA library.
                     cDNA library constructed from mRNA isolated  from C.
                     richardii spores that had developed for 20 hours after
                     their  germination had been initiated by white light."
ORIGIN

Alignment Scores:
Pred. No.:              10.9          Length:        824
Score:                  85.00         Matches:       34
Percent Similarity:     42.45%        Conservative:  25
Best Local Similarity:  24.46%        Mismatches:    32
Query Match:            15.34%        Indels:        48
DB:                     2             Gaps:          6

US-09-837-344-19 (1-107) x BE643089 (1-824)

Qy          3 AlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGlu--- 21
              |||:::|||   |||  ||||||  ||||||          :::|||:::|||
Db        239 GCCGAGACGGTGAAGGCTCTCGAAGCAAAGAAGCGACCATCCATTGTGCTGAGTGAGATT 298

Qy         21 ------------------------------------------------------------ 21

Db        299 GACTCTGATGGCTGGCACGTGATCCTGTCTGAAGGCGGCTTCGTGCTGGATGCGACATCA 358

Qy         22 ------------AspLeuTyrGlyArgLeuGluIle----------------------- 29
                         ||||||    ||||||
Db        359 GAGTGGCCCGCCGATCTCACTGCGCGACTTATCGCTGATCAGCCAGCCGATGCAGTCCAA 418

Qy         30 ------ProAlaIleGluLeuProSerGluAsnGlu---------ArgGlyTyrTyrIle 44
                  ||||||   :::        ||||||||         :::|||   |||:::
Db        419 GTTCCTCCTGCATTCCAGTGGAGTGGAGAGAATGAGCCGGATCAGAAGGGTCGCTACCTC 478

Qy         45 ProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIle 64
              ||||||      :::      |||   :::|||          |||:::::::::   |||

Db        479 CCCCACCTTCGTACTCATCTGCAGCTGCATCGT------TACCGCGAGGCTGAACTGATC 532

Qy         65 SerIleIleGluLysThrAsnArgGluSerIleThrThr--------------AsnVal 79
              |||:::   :::|||   :::      |||::::::                  ::::::
Db        533 TCATTGGATGACAAGAAATCCTTTCTGAGTGTATCCGATCGTCGTCTGCTATTCGACATT 592

Qy         80 GluGlyArgArgAsp---------IleHisLysGlyHisLeuGluGluLysLysAsp 95
              :::|||  .   |||         :::|||||||||::::::      :::::::::
Db        593 AAAGGCACCACCGACGTCGCTGTGGTGCATAAGGGTTACATCGCAGCCCGTGAGGAG 649

RESULT 19
CNS07CXI/c
LOCUS       CNS07CXI                890 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T3 end of clone BD0AA007F05 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL439756
VERSION     AL439756.1  GI:12223168
KEYWORDS    GSS.
SOURCE      Candida tropicalis
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 890)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
   PUBMED   11152891
REFERENCE   3  (bases 1 to 890)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Candida tropicalis"
                     /mol_type="genomic DNA"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA007F05"
                     /clone_lib="BD0AA"
                     /note="end : T3"
     misc_feature    complement(<293. .>844)
                     /note="similar to Saccharomyces cerevisiae ORF YMR250w [
                     similarity to glutamate decarboxylases ]"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:              12            Length:        890
```

```
Score:                 85.00          Matches:       36
Percent Similarity:    37.11%         Conservative:  23
Best Local Similarity: 22.64%         Mismatches:    36
Query Match:           15.34%         Indels:        64
DB:                    11             Gaps:          7

US-09-837-344-19 (1-107) x CNS07CXI (1-890)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||   |||   ||||||                                  |||
Db        781 AGAAAAGTTGATGATAAGAAA------------------------------TTAGAT 755

Qy         21 GluAspLeuTyrGlyArgLeuGlu--------------IleProAlaIle--------- 32
                 ||||||         |||              :::|||   :::
Db        754 TCTAAACTTTACACTATCAATGAAAATTTCAAGCCAGGTTTACCAGTTGTTGCCTTTAGA 695

Qy         33 ----------------------------GluLeuProSerGlu--------------- 37
                                          |||||||||   |||
Db        694 TACTCCAAGATGATTAGGGAAAAATATCCTGAATTGCCACAAGAATTGTTCTCATCATTA 635

Qy         38 ---AsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsn------ 54
                 ::::::||||||   ||||||:::           |||   :::|||
Db        634 ATGAGAAAGAAAGGTTATATCATTCCAAATTATCATTTACCACCAAATGAAAACGATACT 575

Qy         55 ------------------------------------------------ArgGlyAsnSer 58

Db        574 GAATTATTAAGAGTTGTGGTTAGAAATACTTTATCCTTGAACTTGTTGGAAAGTTTAATT 515

Qy         59 ArgAspSerLysGluIle--------------------------SerIleIleGluLys 69
              :::||||||   |||:::                          |||||   |||
Db        514 AAAGATTCAACTGAGGTTGTTGAATTGTTGATCAGAGCCGCTGAATCAATTAAAGAAATT 455

Qy         70 ThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys------ 87
              |||:::::::::::::::   |||:::   :::   |||::::::||||||:::
Db        454 ACATCCAAGCAAAACCTTGAAACTGATGAACAAAAAAGACAAGAGATTCATCGTATTTTA 395

Qy         88 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLys 106
                 :::          |||      ::::::   :::||||||||||:::|||
Db        394 GCAATTATTTCATCTGCTGGCGATGTTGAGGTAAGAGAGAAACAAAAGGAAGAAAAA 338

RESULT 20
DR437990
LOCUS       DR437990                642 bp    mRNA    linear   EST 30-JUN-2005
DEFINITION  ACAC-aaa74h08.g1 Hydra EST UCI 7 Hydra magnipapillata cDNA 5', mRNA
            sequence.
ACCESSION   DR437990
VERSION     DR437990.1  GI:68415861
KEYWORDS    EST.
SOURCE      Hydra magnipapillata
  ORGANISM  Hydra magnipapillata
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
            Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,
            Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
            Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R.,
            Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
  TITLE     WashU Hydra EST Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Hans Bode
            WashU Hydra EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library material provided by Hans Bode & Dirk Lindgens, Univ. of
            Calif., Irvine Library constructed by Dirk Lindgens, Univ. of
            Calif. Irvine Library sequenced by Washington University Genome
            Sequencing Center For information on obtaining a clone please
            contact: Hans Bode (hrbode@uci.edu)
            This trace has been recalled with phred
            original value before phred recall for SL was 87
            original value before phred recall for SR was 902
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Hydra magnipapillata"
                     /mol_type="mRNA"
                     /strain="105"
                     /db_xref="taxon:6085"
                     /lab_host="DH10B"
                     /clone_lib="Hydra EST UCI 7"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; a.
                     1st strand cDNA was primed with a Not I primer-adapter
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3') b.
                     Double-stranded cDNA was ligated to Sal I adapter,
                     digested with Not I and cloned into the pSPORT 1-vector
                     pre-cut with Not I and Sal I. c. The ligation mix was
                     transformed into DH10B cells. d. The cells were grown in
                     SOC = 5g yeast, 20g tryptone, 0.5 g NaCl, 10 mM MgSO4, 10
                     mM MgCl, 0.2% glucose / Liter, (no antibiotic). e. DMSO
                     was added to a final conc. of 10% as a cryoprotectant.and
                     frozen. f. The titre before freezing was determined as ~
                     6000/ 100 ul. Assuming a 10% loss upon freezing, the titre
                     is probably ~ 5400/ 100 ul. g. 6 tubes each containing ~
                     5400 clones/ul [ = total of ~ 32,400] are enclosed. h.
                     The frequency of vectors containing inserts is 95% as
                     determined by digestion check after picking 20 clones,
                     miniprep and subsequent digestion with Not I and Sal I. i.
                     A low level of 32P was used in the cDNA synthesis
                     procedure. The level measured by holding a Geiger Counter
                     next to a tube was background."
ORIGIN

Alignment Scores:
Pred. No.:             9.18           Length:        642
Score:                 84.50          Matches:       30
Percent Similarity:    46.39%         Conservative:  15
Best Local Similarity: 30.93%         Mismatches:    41
Query Match:           15.25%         Indels:        11
DB:                    8              Gaps:          3

US-09-837-344-19 (1-107) x DR437990 (1-642)

Qy          8 AsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu 27
              |||   |||  |||||||||:::  ||||||                      |||
Db        192 AACAACGAAGATAAAAAAGAAAACAATGATATT--------------------AGGAAG 230

Qy         28 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 47
              |||        ||||||          ||||||                       :::
Db        231 GAAGAAAGTCCTATTGAAAAAGAAGAAAGTAACGAAACCAACAATATCAAGGAAGACAAA 290

Qy         48 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 67
                        :::|||   :::   |||      :::   ||||||   |||   |||
Db        291 AAGGAAAGCAACAAAGATAAAAAAGAAAATGATGGTAATGAAAAAGAAGAAAGTAATATT 350

Qy         68 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 87
              ::::::         |||      |||:::|||   |||      :::      :::
Db        351 AAAGAAGAAAGAAATGAA------ACAAGCAACAACGAGGAAGAAAAAAAAGAAAATGAT 404

Qy         88 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGlu 104
              |||:::      ||||||::::::   :::||||||   |||:::   |||
Db        405 GGTAAT------GAAAAAGAAGAAAGTAATATTAAAGAAGAAAGAAATGAA 449

RESULT 21
BH381738
LOCUS       BH381738                693 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-178A21.TF ND-TAM Anopheles gambiae genomic clone
            AG-ND-178A21, genomic survey sequence.
ACCESSION   BH381738
VERSION     BH381738.1  GI:17327880
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
```

```
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
            Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
            Gardner,M.J. and Collins,F.H.
  TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
  JOURNAL   Mol. Genet. Genomics 268 (6), 720-728 (2003)
   PUBMED   12655398
COMMENT     Other_GSSs: AG-ND-178A21.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .693
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-178A21"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:                10.1          Length:          693
Score:                    84.50         Matches:         30
Percent Similarity:       45.79%        Conservative:    19
Best Local Similarity:    28.04%        Mismatches:      43
Query Match:              15.25%        Indels:          15
DB:                       9             Gaps:            4

US-09-837-344-19 (1-107) x BH381738 (1-693)

Qy          3 AlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   |||   |||||||||:::   :::   |||   |||           |||||||
Db        398 GCAGCTACCAATAAGAATCTTCAGGAAGAAATAGAAAAAGGAAAA---TTCCGTGAGGAC 454

Qy         23 LeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyr 42
              ||||||   ||||||   :::      |||   :::|||      ::::::|||
Db        455 CTTTATCACAGACTTTCTGTAATTGAAATTCGTGTTCCTAAGCTTGATGACCGTAAGGAT 514

Qy         43 TyrIlePro-----------HisGlnSerSerLeuProGlnAspAsnArgGlyAsnSer 58
                 ||||||           |||   |||   :::      |||   :::|||   :::
Db        515 GATATTCCTCTTTTAGTGGAGCATTTTTCCAAGATCATTGCGGACGAACAAGGAACTGCT 574

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              :::   :::                |||                |||   |||   |||
Db        575 AAAAAGACTTTCAGCGATGAGTCTATTACAGCATTACAGAATTTTAGCTGGACCGGAAAT 634

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLys------LysAspGly 96
              :::                           |||:::|||:::      |||
Db        635 ATCCGT----------------------GAGTTAAAGGAACGTTGTTGAAAGATTAAT 670

Qy         97 SerIleLysProGluGlnLys 103
```

```
                ::::::::::|||:::   :::
Db        671 AATCTTAGGCCAAAGCCCCGT 691

RESULT 22
CA775281/c
LOCUS       CA775281                805 bp    mRNA    linear   EST 03-DEC-2002
DEFINITION  io87b08.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6133214 3'
            similar to TR:Q14120 Q14120 DBP-5 NUCLEAR PROTEIN. ;, mRNA
            sequence.
ACCESSION   CA775281
VERSION     CA775281.1  GI:26012815
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 805)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .805
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6133214"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

Alignment Scores:
Pred. No.:                12.2          Length:          805
Score:                    84.50         Matches:         32
Percent Similarity:       43.12%        Conservative:    15
Best Local Similarity:    29.36%        Mismatches:      37
Query Match:              15.25%        Indels:          25
DB:                       6             Gaps:            6

US-09-837-344-19 (1-107) x CA775281 (1-805)

Qy          7 LysAsnLeuGluArgLysLysGlu--------------HisGlyAspIleLeuAlaGlu 21
              |||::::::||||||                       |||   :::|||||||
Db        490 AAGGACATGGAACGTCTTACAAGCCTTAGAGCTGGCATTGAAGGACCTTTACTTGCAAGT 431
```

```
Qy         22 AspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGly 41
              |||:::   ||||||                                    :::|||
Db        430 GATGTT---GGACGT----------------------------------GACAGATCT 410

Qy         42 TyrTyrIleProHisGlnSerSerLeuProGln---------AspAsnArgGlyAsnSer 58
                  |||      ||||||:::|||:::         ::::::
Db        409 GCTGCCAGCCCGGTTGTAAGTAGTATGCCAGAAAGAGCTTCAGAGTCTTCTTCAGAGGAA 350

Qy         59 ArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsn 78
              :::|||      ||||||   :::      |||   |||   |||              |||
Db        349 AAAGATGATTATGAAATTTTTGTAAAAGTTAAGGACAATCACGAAAAAAGCAAGAAAAAT 290

Qy         79 ValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
                     ||||||      ||||||      |||:::|||:::|||   |||:::
Db        289 ------AAGANCCGTGAT------AAGGGGGAGAAAGAGAAGAAAAGAGACTCTTCATTA 242

Qy         99 LysProGluGlnLysGluAspLysSer 107
              :::         |||      ||||||
Db        241 AGATCTCGAAGTAAGCGTTCCAAATCT 215

RESULT 23
AG872408/c
LOCUS       AG872408                872 bp    DNA     linear   GSS 03-NOV-2004
DEFINITION  Oryza sativa (indica cultivar-group) genomic DNA, BAC end sequence,
            BAC clone:K0250G12_R, genomic survey sequence.
ACCESSION   AG872408
VERSION     AG872408.1  GI:55338669
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Katagiri,S., Wu,J., Ito,Y., Karasawa,W., Shibata,M., Kanamori,H.,
            Katayose,Y., Namiki,N., Matsumoto,T. and Sasaki,T.
  TITLE     End Sequencing and Chromosomal in silico Mapping of BAC Clones
            Derived from an indica Rice Cultivar, Kasalath
  JOURNAL   Breeding Science 54, 273-279 (2004)
REFERENCE   2  (bases 1 to 872)
  AUTHORS   Sasaki,T., Matsumoto,T. and Wu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2004) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 side of the BAC clone.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Kasalath"
                     /db_xref="taxon:39946"
                     /clone="K0250G12_R"
ORIGIN

Alignment Scores:
Pred. No.:              13.4          Length:        872
Score:                  84.50         Matches:       30
Percent Similarity:     45.87%        Conservative:  20
Best Local Similarity:  27.52%        Mismatches:    40
Query Match:            15.25%        Indels:        19
DB:                     10            Gaps:          4

US-09-837-344-19 (1-107) x AG872408 (1-872)

Qy          6 LysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGly 25
              ||||||   :::||||||::::::|||                |||:::     |||
Db        577 AAGAAGAAAATAGAGAGACAAAGAGAG--------------AAGGAAGAGGANAAAGGA 533

Qy         26 ArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIlePro 45
              :::   :::                                |||||||||   |||
Db        532 CAAAGACAA--------------------AAAGAAAACGAGGGAGGAAGAAAAGAAGAA 494

Qy         46 HisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSer 65
              ::::::   :::      |||      :::|||      :::   :::|||:::
Db        493 AACGAAAGAGCAAAAAAGCAAAAAAAGAAAGGAAGAGGGAAGAAGACAAAGAAGCAAAGA 434

Qy         66 IleIleGluLys------------ThrAsnArgGluSerIleThrThrAsnValGluGly 81
              ||||||               :::|||::::::                :::|||
Db        433 GAAAGAGAGAAACGAAAGAAGGAAAAAAGCAGAAAAGCGGACAAACAAGAACAAGACGGG 374

Qy         82 ArgArgAsp---------IleHisLysGlyHisLeuGluGluLysLysAspGlySerIle 98
              ||||||:::          |||   |||   :::   |||||||||
Db        373 AGGAGAGAAGAAAGAAACACGCGAAAAAAACACCCCAAAAGGAAGAAGGACGAAAGAGCA 314

Qy         99 LysProGluGlnLysGluAspLysSer 107
              |||   |||||||||      |||:::
Db        313 AAAAGAGAACAAAAAAACAGAAAAACA 287

RESULT 24
AG453925
LOCUS       AG453925                1513 bp    DNA     linear   GSS 22-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-339L07.TJ, genomic survey
            sequence.
ACCESSION   AG453925
VERSION     AG453925.1  GI:48155439
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1513)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1513
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-339L07.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                26.7          Length:        1513
Score:                    84.50         Matches:       32
Percent Similarity:       44.44%        Conservative:  28
Best Local Similarity:    23.70%        Mismatches:    42
Query Match:              15.25%        Indels:        33
DB:                       10            Gaps:          6

US-09-837-344-19 (1-107) x AG453925 (1-1513)

Qy          2 LysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGlu 21
              :::   :::   ||||||     |||:::|||||||||   |||:::|||:::   :::
Db        448 AGAAAAGAAAAAAAGAAAAAAGAGGAGAAGAAAAAGGAGAAAGGAAATATTGTGAAGAAA 507

Qy         22 AspLeuTyrGlyArgLeuGlu-----------IleProAlaIleGluLeuProSerGlu 37
                         ||||||::::::           :::   :::::::::          |||
Db        508 ---------GGAAGAATAAAAAGGAGGAAGGGGGTAGGAAGTATGAAGAAGAGATGAGAA 558

Qy         38 AsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu-----------------Pro 51
                 |||||||||                 :::   :::
Db        559 AAAGAGAGAGGAAGAAGGAAAAAAGAGAAAGAAAAAGTGGTATTTATTGATAAAGAAAAA 618

Qy         52 GlnAspAsnArgGlyAsnSerArgAspSerLysGlu----------------------- 63
              ::::::                 ||||||:::||||||
Db        619 GAAGAAGAAGAGGAAGGAGAGAGGGATAATAAGGAGGAAAATAATTTATAAAAAAAAAAA 678

Qy         64 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArg--- 82
              ::::::::::::::       |||                |||   |||   :::
Db        679 AACAATGTATTAAAAAGAATAAAGAGGAGGGAAAAGAATAAAAATAAAGAAGAAAAAAAT 738

Qy         83 -----------------ArgAspIleHisLysGlyHisLeuGluGluLys--------- 93
                              :::         ||||||   |||:::::|||
Db        739 AAAAAGTAAGAAAGAGGTAAGAAGAAAAGGAAAGGGGAGTTAAAAGATAAGAAAGAGGAG 798

Qy         94 ------LysAspGlySerIleLysProGluGlnLysGluAspLys 106
                   |||||||||       |||   |||:::   |||:::|||
Db        799 TAAAATAAAGATGGAGAGAAGAAGAAAGAGGAAGGGGAAGAAAAG 843

RESULT 25
DE006900/c
LOCUS       DE006900                606 bp    DNA     linear   GSS 06-APR-2005
DEFINITION  Branchiostoma floridae DNA, clone: CH302-012D18.R, genomic survey
            sequence.
ACCESSION   DE006900
VERSION     DE006900.1  GI:62249272
KEYWORDS    GSS.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1
  AUTHORS   Fujiyama,A., Toyoda,A., Hattori,M. and Sakaki,Y.
  TITLE     BAC end sequences of CH0RI-302 Amphioxus Library
  JOURNAL   Published Only in Database (2005)
REFERENCE   2  (bases 1 to 606)
  AUTHORS   Fujiyama,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2005) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:afujiyam@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     PRIMERS
                Sequencing : TJ
            LIBRARY
                Vector     : pTARBAC2.1
                R.Site 1   : EcoRI
                R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Branchiostoma floridae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7739"
                     /clone="CH302-012D18.R"
                     /sex="male"
                     /tissue_type="sperm"
                     /clone_lib="CH0RI0302 Amphioxus genomic BAC library"
                     /note="common name:amphioxus"
ORIGIN

Alignment Scores:
Pred. No.:                9.81          Length:        606
Score:                    84.00         Matches:       24
Percent Similarity:       46.59%        Conservative:  17
Best Local Similarity:    27.27%        Mismatches:    25
Query Match:              15.16%        Indels:        22
DB:                       11            Gaps:          4

US-09-837-344-19 (1-107) x DE006900 (1-606)

Qy         12 LysLysGluHis---------GlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGlu 28
              ||||||   |||          ||||||::::::
Db        226 AAAAAACGACACAGCACTCTTGGAGACATGATTTGT---------------------- 191

Qy         29 IleProAlaIleGluLeuProSer---------------GluAsnGluArgGlyTyrTyr 43
              :::|||   |||   :::   |||                ::::::   |||||||||
Db        190 CTCCCAACCATATAAATATACAGTGTGCAGTGAAATACACAAGATGTGCGCGGATATGTT 131

Qy         44 IleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGlu 63
              :::|||   ::::::   ||||||:::       |||       :::||||||||||||
Db        130 TTGCCACAAGAGACAGGCCTTCCTCGTTCAACACGTTCAAGGACAAGAGATTCAAAGATC 71

Qy         64 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg 83
              :::         :::         :::   :::   :::|||
Db         70 CTGAAGGAGGCCGACTGACCTGTGAAGGCCACATCCAGCACAAAG-------------- 26

Qy         84 AspIleHisLysGlyHisLeuGlu 91
                  :::|||   ||||||   |||
Db         25 ---CTACATGTAGGTCACTTTGAG 5

RESULT 26
DN674862
LOCUS       DN674862                1439 bp    mRNA    linear   EST 29-MAR-2005
DEFINITION  CFW83-A07.y1d-s SHGC-CFW Gasterosteus aculeatus cDNA clone
            CFW83-A07 5', mRNA sequence.
ACCESSION   DN674862
VERSION     DN674862.1  GI:61995041
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1439)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Ave, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 83
            High quality sequence start: 49
            High quality sequence stop: 902.
FEATURES             Location/Qualifiers
     source          1. .1439
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Conner Creek sticklebacks, WA USA"
                     /db_xref="taxon:69293"
```

```
                     /clone="CFW83-A07"
                     /sex="mixed male and female"
                     /tissue_type="gills"
                     /dev_stage="adult"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="SHGC-CFW"
                     /note="Vector: Express 1; Total and poly A+ RNA was
                     isolated from the indicated stickleback tissue, and a cDNA
                     library was constructed in the Express 1 plasmid vector by
                     Open Biosystems. First strand cDNA synthesis was primed
                     with an 54 bp linker primer containing an oligodT sequence
                     preceeded by a synthetic NotI site (first strand primer:
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following
                     second strand synthesis, cDNAs were made blunt at the end
                     corresponding to the original 5 prime end of mRNA, and
                     cloned directionally into the NotI and EcoRV sites of
                     Express 1. Note that the EcoRV site is typically destroyed
                     in the blunt end cloning, leaving a junction of the form
                     'xxxATC' (where is ATC is the second half of the EcoRV
                     site, and xxx is derived from the cDNA sequence). A map of
                     the Express 1 vector is available at:
                     http://www.openbiosystems.com/cdna_library_construction_fa
                     q.php#8 The primary library was transformed and amplified
                     in DH10B (T1 phage resistant) bacteria. Clones available
                     from Open Biosystems:
                     http://www.openbiosystems.com/stickleback"
ORIGIN

Alignment Scores:
Pred. No.:              28.8          Length:        1439
Score:                  84.00         Matches:       35
Percent Similarity:     44.54%        Conservative:  18
Best Local Similarity:  29.41%        Mismatches:    46
Query Match:            15.16%        Indels:        20
DB:                     8             Gaps:          6

US-09-837-344-19 (1-107) x DN674862 (1-1439)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              |||:::   |||     |||       :::   ||||||   ||||||   :::||| 
Db        370 AGGGAGCGTGATGCCATCAAGGAG---TCCAAGAGCAAGGAGAAAGGAGACAGAATCGCA 426

Qy         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer---GluAsnGlu 39
                 :::   |||   |||:::   |||          |||   |||   |||:::|||
Db        427 ------GTGGCTGGGTCCCTAAAGTCACCGGTTCCCAGATTGGAATCAGCCGAATCTGAG 480

Qy         40 Arg-----------------------GlyTyrTyrIleProHisGlnSerSerLeuPro 51
              |||                      :::     |||      ||||||
Db        481 AGGGATCACAAAAGACGAAAGCTCGACAGTCACTCTTCTCCGTCCCACTCCTCGTCTGTT 540

Qy         52 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 71
              :::|||         ||||||   :::   ||||||   :::              |||
Db        541 AAGGAC---------AATAGCAACGAACCCAAGGAGTCAACATCCAAGCACCACGCCAAC 591

Qy         72 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys----------- 87
                 |||:::|||   :::           |||::::::   |||
Db        592 TACAATTCAGTTACCCGGTCGAAGAGCAGAGAGAGGGAGATGGAGAAGAAAGATTCAGAC 651

Qy         88 ---GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAsp 105
                    :::||||||||||:::         :::   |||:::|||   |||
Db        652 AACACACGATCCAAAGAGAAGAAGGAAGAAAAGGATCGTAAAGAAAGGAAGAGAGAT 708

RESULT 27
BX320026
LOCUS       BX320026                602 bp    mRNA    linear   EST 20-MAY-2004
DEFINITION  BX320026 tcay Oncorhynchus mykiss cDNA clone tcay0036b.f.21 3prim,
            mRNA sequence.
ACCESSION   BX320026
VERSION     BX320026.2  GI:40309517
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Govoroun,M., Guiguen,Y. and Le Gac,F.
  TITLE     Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
  JOURNAL   Unpublished (2003)
COMMENT     On Apr 7, 2003 this sequence version replaced gi:29600671.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0036  row: f  column: 21
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcay0036b.f.21"
                     /tissue_type="adipose tissue, blood, brain,
                     differentiating gonads, gills, interrenal, intestine,
                     kidney, liver, muscle, ovary, pituitary, testis"
                     /dev_stage="from embryos to adults"
                     /lab_host="DH10B"
                     /clone_lib="tcay"
                     /note="Vector: pT7T3D-pac; Rainbow trout multi-tissues -
                     normalized + 1 subtraction (tcay) ; Clone distribution :
                     AGENAE Resource centre. Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, France, +33 (0) 1.34.65.28.02, +33
                     (0) 1.34.65.22.73"
ORIGIN

Alignment Scores:
Pred. No.:              11.2          Length:        602
Score:                  83.50         Matches:       31
Percent Similarity:     46.02%        Conservative:  21
Best Local Similarity:  27.43%        Mismatches:    48
Query Match:            15.07%        Indels:        13
DB:                     5             Gaps:          5

US-09-837-344-19 (1-107) x BX320026 (1-602)

Qy          3 AlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAsp 22
              |||   :::         :::||||||   |||      |||   :::   |||
Db        104 GCCTCGTCTCTGGCCGAGGTGGAGAGGTTGAAG------GGGATGCTGCAGGCTGGTCAG 157

Qy         23 LeuTyrGlyArg-----------------LeuGluIleProAlaIleGluLeuProSer 36
              :::   ||||||                 ||||||:::      :::|||
Db        158 ATCCCCGGACGGGAGGTCAGACAAGTCCCCCTGGAAATGGTGGAGGTGGAA---GAGGAA 214

Qy         37 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 56
              |||   |||   |||       :::   |||                |||       |||
Db        215 GAGGAGGAGAACGGG------GTTGTACATATGCAACGAGAGATACAGACAGAGGAAGGA 268

Qy         57 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 76
              |||      :::      |||      ::::::|||:::         :::
Db        269 AATGGAGTGGAGGAGATGGAAGAGGATGTGGTGGAGGAGGTACAGGGAAAGGTGGAGGAG 328

Qy         77 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 96
                 :::|||:::         ::::::   |||      :::||||||||:::::|||
Db        329 GTAGATGTCAAGGAGGTGGAGGAGGTAGAAAAGGAAAAAGTGGAGGAGAAAGAAGAGGGG 388

Qy         97 ------SerIleLysProGluGlnLysGluAspLysSer 107
                       |||   |||   :::||||||:::|||
Db        389 GAGGAGGGTGAAAAACAAGAGTCTGAGGAAGATGAGTCT 427
```

```
RESULT 28
CW816235/c
LOCUS       CW816235               621 bp    DNA     linear   GSS 20-JAN-2005
DEFINITION  pinea_25942 Phytophthora infestans T30-4 Phytophthora infestans
            genomic, genomic survey sequence.
ACCESSION   CW816235
VERSION     CW816235.1  GI:57990652
KEYWORDS    GSS.
SOURCE      Phytophthora infestans (potato late blight agent)
  ORGANISM  Phytophthora infestans
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Randall,T.A., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C.,
            Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,E., Gaffney,T.,
            Law,M., Testa,A., Torto-Alalibo,T., Zhang,M., Zheng,L., Mueller,E.,
            Windass,J., Binder,A., Birch,P., Gisi,U., Govers,F., Gow,N.A.,
            Mauch,F., van West,P., Waugh,M.E., Yu,J., Boller,T., Kamoun,S.,
            Lam,S.T. and Judelson,H.S.
  TITLE     Large-scale gene discovery in the oomycete Phytophthora infestans
            reveals likely components of phytopathogenicity shared with true
            fungi
  JOURNAL   Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
   PUBMED   15782637
COMMENT     Contact: Judelson HS
            Department of Plant Pathology
            University of California
            PLPA, Riverside, CA 92521, USA
            Tel: 9518274199
            Fax: 9518274294
            Email: howard.judelson@ucr.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Phytophthora infestans"
                     /mol_type="genomic DNA"
                     /isolate="T30-4"
                     /db_xref="taxon:4787"
                     /sex="A1"
                     /tissue_type="total mycelium"
                     /clone_lib="Phytophthora infestans T30-4"
                     /note="Vector: pUC18"
ORIGIN

Alignment Scores:
Pred. No.:                11.6          Length:          621
Score:                    83.50         Matches:         27
Percent Similarity:       49.43%        Conservative:    16
Best Local Similarity:    31.03%        Mismatches:      36
Query Match:              15.07%        Indels:          8
DB:                       10            Gaps:            3

US-09-837-344-19 (1-107) x CW816235 (1-621)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHis-----------Gly 16
              |||   |||      :::      :::      |||:::   |||           |||
Db        543 CGTCCAGCGCGTGACGAAAACTATGTCGGAGGAAAGAGAGCGCACCGTGACAATCGTGGT 484

Qy         17 AspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIle-GluLeuProSe 36
                 :::   |||:::   |||   |||   :::   ||||||:::::: ||||||
Db        483 TCCGTCTCTGCGGACGCGCTCGTTGGTCCGATACGGATACCAAGCGTCCGAGTTGACAAA 424

Qy         36 rGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGl 56
                |||      |||   |||      ||||||:::|||   :::|||
Db        423 GTCTGAGACCAGAACTCCGTACCCACCTTTTCCGTCGTCCATTCCGTCGGAAAACCTCGC 364

Qy         56 yAsnSerArg------AspSerLysGluIleSerIle---IleGluLysThrAsnArgGl 73
              ||||||         ||||||      :::|||      :::      |||:::|||
Db        363 AAACAGCAGTGTATTGGACTCCACGGTAGTGTCGGCCGAAGTCGCAGTAACCAGCCGAAG 304

Qy         73 uSerIleThrThrAsnVal 79
                 :::   ::::::|||
Db        303 TTGTGTGAACTCGGACGTC 285

RESULT 29
BW317679/c
LOCUS       BW317679               633 bp    mRNA    linear   EST 11-NOV-2002
DEFINITION  BW317679 Nori Satoh unpublished cDNA library, heart Ciona
            intestinalis cDNA clone ciht036e05 5', mRNA sequence.
ACCESSION   BW317679
VERSION     BW317679.1  GI:24898363
KEYWORDS    EST.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona intestinalis (2002c)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="ciht036e05"
                     /tissue_type="heart"
                     /clone_lib="Nori Satoh unpublished cDNA library, heart"
ORIGIN

Alignment Scores:
Pred. No.:                11.9          Length:          633
Score:                    83.50         Matches:         39
Percent Similarity:       41.26%        Conservative:    20
Best Local Similarity:    27.27%        Mismatches:      37
Query Match:              15.07%        Indels:          47
DB:                       5             Gaps:            8

US-09-837-344-19 (1-107) x BW317679 (1-633)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ||||||:::      :::|||   |||   |||   |||   :::   :::||||||
Db        519 CGAAAAAGTCATCACGAAAAACAGCTTTATCGACTAAAATCAAACGAAGAAATATTAAAA 460

Qy         21 GluAspLeu---TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGlu 39
              :::|||:::   :::|||              ||||||              |||
Db        459 GACGACATAAGTTTCGGCGATCGACTCTCCGACGCCATCTCGCGGCAGTGGAATAAT--- 403

Qy         40 ArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArg 59
                                        |||   :::      ||||||         |||
Db        402 ----------------------TCGGAAATGGCAAATGATAACGGAGAGGTTTCATCT 367

Qy         60 AspSerLys----------------------GluIleSerIleIleGluLysThrAsn 71
                ||:::                        ||||||   :::   ||||||   |||
Db        366 AGTTCTGAATCAGATGATAGCGACATAGAGACAGAAATAGAAGTCAAAGAAAAGGTAAAC 307

Qy         72 ArgGluSerIleThrThr--------------------AsnValGluGlyArgArgAsp 84
                :::||||||                          |||         :::||||||
Db        306 GTGAATACGATAACTGTGGTGACAAAGACGGAACAGGAGAACCAAGCAAACAAACGCGAT 247

Qy         85 IleHis-----------------LysGly---------------------------- 88
                |||                  ||||||
Db        246 GATCATGAAGATGTGGAAACAGTTAAAGGGGAATACAAAGTAGAGGAGGAGGATGTTTGT 187

Qy         89 ------HisLeuGluGluLys------LysAspGlySerIleLysProGluGlnLysGlu 104
                  |||::::::   |||      |||:::||||||:::      |||:::|||:::
Db        186 AGAGTTCACATCAAAGCTAAATATATAAAGGAAGGATCGCTG------GAAGAAAAACAA 133
```

```
Qy        105 AspLysSer 107
              :::|||:::
Db        132 GAAAAAACT 124

RESULT 30
CO131269/c
LOCUS       CO131269                684 bp    mRNA    linear   EST 17-JUN-2004
DEFINITION  GR__Eb43H18.f GR__Eb Gossypium raimondii cDNA clone GR__Eb43H18 5',
            mRNA sequence.
ACCESSION   CO131269
VERSION     CO131269.1  GI:48880247
KEYWORDS    EST.
SOURCE      Gossypium raimondii
  ORGANISM  Gossypium raimondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
            Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
            Wing,R.A.
  TITLE     Global assembly of Cotton ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 43  row: H  column: 18.
FEATURES             Location/Qualifiers
     source          1. .684
                     /organism="Gossypium raimondii"
                     /mol_type="mRNA"
                     /db_xref="taxon:29730"
                     /clone="GR__Eb43H18"
                     /tissue_type="floral"
                     /dev_stage="-3 to +3 DPA"
                     /lab_host="DH10B"
                     /clone_lib="GR__Eb"
                     /note="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:
                     EcoRV; Library made by Invitrogen with RNA supplied by
                     Wendle lab. Directional cloned into NotI-EV. Colonies
                     plated/picked by AGI. More glycerol clones held in -80."
ORIGIN

Alignment Scores:
Pred. No.:              13.1          Length:         684
Score:                  83.50         Matches:        30
Percent Similarity:     45.45%        Conservative:   20
Best Local Similarity:  27.27%        Mismatches:     53
Query Match:            15.07%        Indels:         7
DB:                     7             Gaps:           3

US-09-837-344-19 (1-107) x CO131269 (1-684)

Qy          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20
              ::::::|||      ||||||      ::::::||||||:::      |||
Db        543 AAAGAGGCAAAGGAGAAGAAGAAAGAAAAGAAGAAGAAGGATGAGGTAGATGAAGATGAA 484

Qy         21 GluAspLeuTyrGlyArgLeuGlu------------IleProAlaIleGluLeuProSer 36
              |||         |||:::   |||             |||   |||
Db        483 GAATTAGAAGAGGGAAAGAAAGAAAAGAAGAAAGACAAAGAGGCGAAGGAGAAGAAGAAA 424

Qy         37 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 56
              |||   ::::::         :::               |||   :::   |||:::
Db        423 GAAAAGAAGAAGAAGGATGAGGTAGAAGAAGCTGAAGAATTAGAAGAGGGAAATAAAGAA 364

Qy         57 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 76
                 :::|||   ||||||            ||||||          :::
Db        363 AAGAAGAAAGAC---AAAGAGGCGAAG------GAGAAGAAGAAAGAAAAGAAGAAGAAG 313

Qy         77 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 96
                    ||||||         ::::::   :::|||:::   |||:::|||||||||
Db        312 GATGAGGTAGAAGAAGCTGAAGAATTAGAAGAGGGAAATAAAGAAAAGAAGAAAGACAAA 253

Qy         97 SerIleLysProGluGlnLysGluAspLys 106
                 |||   ::::::||||||   |||
Db        252 GAGGCAAAGGAGAAGAAGAAAGAAAAAAG 223


Search completed: December 20, 2005, 05:09:23
Job time : 3252 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2005, 21:38:01 ; Search time 229 Seconds
(without alignments)
329.657 Million cell updates/sec

Title: US-09-837-344-19
Perfect score: 554
Sequence: 1 RKADTKKNLERKKEHGDILA.....GHLEEKKDGSIKPEQKEDKS 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	554	100.0	280	2	Q25844_PLAFA	Q25844 plasmodium
2	554	100.0	280	2	Q25849_PLAFA	Q25849 plasmodium
3	554	100.0	280	2	Q25850_PLAFA	Q25850 plasmodium
4	554	100.0	493	2	Q25886_PLAFA	Q25886 plasmodium
5	553	99.8	280	2	Q25843_PLAFA	Q25843 plasmodium
6	553	99.8	280	2	Q25846_PLAFA	Q25846 plasmodium
7	553	99.8	280	2	Q25847_PLAFA	Q25847 plasmodium
8	553	99.8	280	2	Q25851_PLAFA	Q25851 plasmodium
9	553	99.8	1909	2	Q25893_PLAFA	Q25893 plasmodium
10	550	99.3	280	2	Q25848_PLAFA	Q25848 plasmodium
11	550	99.3	280	2	Q25852_PLAFA	Q25852 plasmodium
12	549	99.1	280	2	Q25845_PLAFA	Q25845 plasmodium
13	549	99.1	280	2	Q25853_PLAFA	Q25853 plasmodium
14	549	99.1	280	2	Q25854_PLAFA	Q25854 plasmodium
15	549	99.1	280	2	Q25855_PLAFA	Q25855 plasmodium
16	549	99.1	280	2	Q25888_PLAFA	Q25888 plasmodium
17	549	99.1	280	2	Q25889_PLAFA	Q25889 plasmodium
18	549	99.1	280	2	Q25900_PLAFA	Q25900 plasmodium
19	549	99.1	280	2	Q27243_PLAFA	Q27243 plasmodium
20	548	98.9	1596	2	Q8IJ44_PLAF7	Q8ij44 plasmodium
21	469	84.7	264	2	Q9GTX5_PLAFA	Q9gtx5 plasmodium
22	224	40.4	42	2	Q8MZK9_PLAFA	Q8mzk9 plasmodium
23	97	17.5	1130	2	Q59YV6_CANAL	Q59yv6 candida alb
24	91	16.4	210	2	O66502_AQUAE	O66502 aquifex aeo
25	86.5	15.6	354	2	Q7RQI4_PLAYO	Q7rqi4 plasmodium
26	85	15.3	710	2	Q7RJN5_PLAYO	Q7rjn5 plasmodium
27	85	15.3	3812	2	Q4RLC8_TETNG	Q4rlc8 tetraodon n
28	84.5	15.3	429	2	Q6R1M2_PLAFA	Q6r1m2 plasmodium
29	84	15.2	457	2	Q8R5C4_MOUSE	Q8r5c4 mus musculu
30	84	15.2	596	2	Q56PC1_CALJA	Q56pc1 callithrix
31	84	15.2	689	2	Q7R9G0_PLAYO	Q7r9g0 plasmodium
32	83.5	15.1	603	2	Q9AHN3_PASMU	Q9ahn3 pasteurella
33	82.5	14.9	531	2	Q6AJU9_DESPS	Q6aju9 desulfotale
34	82.5	14.9	650	2	Q7RCN9_PLAYO	Q7rcn9 plasmodium
35	82	14.8	202	2	Q8GW95_ARATH	Q8gw95 arabidopsis
36	82	14.8	294	2	Q895S6_CLOTE	Q895s6 clostridium
37	82	14.8	571	2	Q4YTI4_PLABE	Q4yti4 plasmodium
38	82	14.8	786	2	Q9FMJ2_ARATH	Q9fmj2 arabidopsis
39	82	14.8	829	2	Q8L7V3_ARATH	Q8l7v3 arabidopsis
40	81.5	14.7	442	2	O66551_AQUAE	O66551 aquifex aeo
41	81.5	14.7	485	2	Q75EU2_ASHGO	Q75eu2 ashbya goss
42	81.5	14.7	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
43	81	14.6	286	2	Q5WE27_BACSK	Q5we27 bacillus cl
44	81	14.6	568	2	Q55EZ8_DICDI	Q55ez8 dictyosteli
45	81	14.6	668	2	Q4UIN9_THEAN	Q4uin9 theileria a

ALIGNMENTS

RESULT 1
Q25844_PLAFA
ID Q25844_PLAFA PRELIMINARY; PRT; 280 AA.
AC Q25844;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; L40885; AAB59231.1; -; Genomic_DNA.
FT NON_TER 1 1
SQ SEQUENCE 280 AA; 32940 MW; E9708E3CFFAEA9CF CRC64;

Query Match 100.0%; Score 554; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.1e-41;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy 61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
|||||||||||||||||||||||||||||||||||||||||||||
Db 61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107

RESULT 2
Q25849_PLAFA
ID Q25849_PLAFA PRELIMINARY; PRT; 280 AA.
AC Q25849;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,

```
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40890; AAC41600.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32966 MW;  E96D255154DEA9CF CRC64;

  Query Match             100.0%;  Score 554;  DB 2;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-41;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 3
Q25850_PLAFA
ID   Q25850_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25850;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40891; AAC41601.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32989 MW;  E96F812CFFAEBED8 CRC64;

  Query Match             100.0%;  Score 554;  DB 2;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-41;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 4
Q25886_PLAFA
ID   Q25886_PLAFA PRELIMINARY;      PRT;   493 AA.
AC   Q25886;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=94267224; PubMed=7515922;
RA   Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA   Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA   Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT   "Plasmodium falciparum liver stage antigen-1 is well conserved and
RT   contains potent B and T cell determinants.";
RL   J. Immunol. 153:190-204(1994).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA   Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA   Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT   "The Plasmodium falciparum liver stage antigen LSA-1 is well conserved
RT   and harbors major B- and T-cell epitopes.";
RL   J. Immunol. 0:0-0(1994).
DR   EMBL; Z30320; CAA82975.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   493 AA;  59062 MW;  FE252CC44F46996E CRC64;

  Query Match             100.0%;  Score 554;  DB 2;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-41;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 273

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        274 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 320


RESULT 5
Q25843_PLAFA
ID   Q25843_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25843;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40884; AAB59230.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32926 MW;  467080F32FAEAD33 CRC64;

  Query Match             99.8%;  Score 553;  DB 2;  Length 280;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-41;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 6
Q25846_PLAFA
```

```
ID   Q25846_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40887; AAC41597.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32882 MW;  46763641C81AFC33 CRC64;

  Query Match             99.8%;  Score 553;  DB 2;  Length 280;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-41;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 7
Q25847_PLAFA
ID   Q25847_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25847;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40888; AAC41598.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32944 MW;  467085E23BABBC27 CRC64;

  Query Match             99.8%;  Score 553;  DB 2;  Length 280;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-41;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 8
Q25851_PLAFA
ID   Q25851_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40893; AAC41603.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32927 MW;  48706EFDCF40ADD3 CRC64;

  Query Match             99.8%;  Score 553;  DB 2;  Length 280;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-41;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 9
Q25893_PLAFA
ID   Q25893_PLAFA PRELIMINARY;      PRT;  1909 AA.
AC   Q25893;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Liver stage antigen.
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NF54;
RX   MEDLINE=92107224; PubMed=1840628; DOI=10.1016/0166-6851(91)90117-O;
RA   Zhu J., Hollingdale M.R.;
RT   "Structure of Plasmodium falciparum liver stage antigen-1.";
RL   Mol. Biochem. Parasitol. 48:223-226(1991).
DR   EMBL; X56203; CAA39663.1; -; Genomic_DNA.
DR   PIR; S24597; A45592.
DR   HSSP; O15813; 1D7M.
SQ   SEQUENCE   1909 AA;  230153 MW;  8B861A524C797217 CRC64;

  Query Match             99.8%;  Score 553;  DB 2;  Length 1909;
  Best Local Similarity   99.1%;  Pred. No. 1.3e-40;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db       1630 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 1689

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1690 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 1736


RESULT 10
Q25848_PLAFA
ID   Q25848_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40889; AAC41599.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32939 MW;  E77080385FAA0D61 CRC64;

  Query Match             99.3%;  Score 550;  DB 2;  Length 280;
  Best Local Similarity   99.1%;  Pred. No. 2.5e-41;
  Matches  106;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPKQKEDKS 107


RESULT 11
Q25852_PLAFA
ID   Q25852_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25852;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40834; AAB59233.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32927 MW;  E6648F85FBBAA245 CRC64;

  Query Match             99.3%;  Score 550;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 2.5e-41;
  Matches  105;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||:|||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEEKEDKS 107


RESULT 12
Q25845_PLAFA
ID   Q25845_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25845;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40886; AAB59232.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32908 MW;  FDEE929F34DCA7E9 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||:|||||||||||| ||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 13
Q25853_PLAFA
ID   Q25853_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25853;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40835; AAB59234.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32938 MW;  46751C45F4DCBD33 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 14
Q25854_PLAFA
ID   Q25854_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25854;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40836; AAC41604.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32966 MW;  61851C5AD6E1A211 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 15
Q25855_PLAFA
ID   Q25855_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25855;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40837; AAC41605.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32866 MW;  73164A709789DD31 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 16
Q25888_PLAFA
ID   Q25888_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25888;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40908; AAB59228.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32936 MW;  FDF9A29104C5A7E9 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              ||||||||||||||||||:||||||||||| ||||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 17
Q25889_PLAFA
ID   Q25889_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40909; AAB59229.1; -; Genomic_DNA.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   280 AA;  32850 MW;  FDEE929F34DF68E9 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||| |||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 18
Q25900_PLAFA
ID   Q25900_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25900;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40947; AAC41592.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32911 MW;  F8850CABF6136212 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||| |||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 19
Q27243_PLAFA
ID   Q27243_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q27243;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40910; AAC41582.1; -; Genomic_DNA.
DR   EMBL; L40892; AAC41602.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32881 MW;  105849F296958C78 CRC64;

  Query Match             99.1%;  Score 549;  DB 2;  Length 280;
  Best Local Similarity   98.1%;  Pred. No. 3.1e-41;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:|||||||||||| |||||||||||||||||||||||||||||
Db          1 RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYIPHQSSLPQDNRGNSRD 60

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107


RESULT 20
Q8IJ44_PLAF7
ID   Q8IJ44_PLAF7 PRELIMINARY;      PRT;  1596 AA.
AC   Q8IJ44;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Liver stage antigen, putative.
GN   ORFNames=PF10_0356;
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.G.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014834; AAN35553.1; -; Genomic_DNA.
SQ   SEQUENCE   1596 AA;  191691 MW;  C0B3A2CC0F5D60AC CRC64;

  Query Match             98.9%;  Score 548;  DB 2;  Length 1596;
  Best Local Similarity   98.1%;  Pred. No. 3e-40;
  Matches  105;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 60
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       1317 RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRD 1376

Qy         61 SKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||| |||||||||||||||||||||||||||||||||||||||
Db       1377 SKEISIIENTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 1423


RESULT 21
Q9GTX5_PLAFA
ID   Q9GTX5_PLAFA PRELIMINARY;      PRT;   264 AA.
AC   Q9GTX5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Xu J., Chen S.Y.;
RT   "Cloning and sequence analysis of 3' region of liver stage antigen-1
RT   gene of Plasmodium falciparum isolate FCC1/HN.";
RL   Zhongshan Yi Ke Da Xue Xue Bao 22:410-413(2001).
DR   EMBL; AF246996; AAG12324.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   264 AA;  31006 MW;  1714D653E8D9D1D7 CRC64;

  Query Match             84.7%;  Score 469;  DB 2;  Length 264;
  Best Local Similarity   98.9%;  Pred. No. 4.5e-34;
  Matches   90;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         17 DILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESIT 76
              |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESIT 60

Qy         77 TNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 107
              |||||||||||||||||||||||||||||||
Db         61 TNVEGRRDIHKGHLEEKKDGSIKPEQKEDKS 91


RESULT 22
Q8MZK9_PLAFA
ID   Q8MZK9_PLAFA PRELIMINARY;      PRT;    42 AA.
AC   Q8MZK9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage antigen (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=21996440; PubMed=12000842; DOI=10.1093/nar/30.10.2224;
RA   Blair P.L., Witney A., Haynes J.D., Moch J.K., Carucci D.J.,
RA   Adams J.H.;
RT   "Transcripts of developmentally regulated Plasmodium falciparum genes
RT   quantified by real-time RT-PCR.";
RL   Nucleic Acids Res. 30:2224-2231(2002).
DR   EMBL; AY090552; AAM09070.1; -; Genomic_DNA.
FT   NON_TER       1      1
FT   NON_TER      42     42
SQ   SEQUENCE   42 AA;  4750 MW;  392CB3001EACD7AD CRC64;

  Query Match             40.4%;  Score 224;  DB 2;  Length 42;
  Best Local Similarity   97.6%;  Pred. No. 5.5e-13;
  Matches   41;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 HGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRG 56
              |||:|||||||||||||||||||||||||||||||||||||||
Db          1 HGDVLAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRG 42


RESULT 23
Q59YV6_CANAL
ID   Q59YV6_CANAL PRELIMINARY;      PRT;  1130 AA.
AC   Q59YV6;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=CaO19.6967;
OS   Candida albicans SC5314.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=237561;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=SC5314;
RX   PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA   Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,
RA   Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA   Davis R.W., Scherer S.;
RT   "The diploid genome sequence of Candida albicans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=SC5314;
RA   Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
RA   Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,
RA   Jones T., Scherer S., Agabian N.;
RT   "Annotation of the Genome of Candida albicans.";
RL   Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AACQ01000098; EAK95626.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1130 AA;  128638 MW;  13E33CE2054587E9 CRC64;

  Query Match             17.5%;  Score 97;  DB 2;  Length 1130;
  Best Local Similarity   24.7%;  Pred. No. 6.9;
  Matches   37;  Conservative   29;  Mismatches   36;  Indels   48;  Gaps    9;

Qy          1 RKADTKKNLERKKEHGDI---------LAEDLYGRLEIPA-------IELPSENERGYYI 44
              :: ||  |:: |||   |         : |||  ::| |:       :|:||::|  : |
Db        883 KEVDTDSNIQEKKELFQIKEPVKDEKQIKEDLV-KVETPSKDEEPAKVEVPSKDETPFKI 941

Qy         45 PHQSSLPQDNRGNSRDSK----------EISIIEKTNRESI----------TTNVEGRRD 84
                :  |   || | ::| |          :  :: | ::: |            : ||::|
Db        942 QNIDS-TTDN-GKAKDEKDNVSKVDSNVDTKVVTKDDQDGIKKVEEKPVEKKDDDEGKKD 999

Qy         85 -----IHKGHLEEKKDGSIKPE----QKED 105
                  :|| : || ||  :       ||:|
Db       1000 NEKVTVHKDNEEEDKDKGVNEAKDKVQKDD 1029


RESULT 24
O66502_AQUAE
ID   O66502_AQUAE PRELIMINARY;      PRT;   210 AA.
AC   O66502;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Transcriptional regulator (H-T-H).
GN   Name=hth; OrderedLocusNames=AQ_093;
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000673; AAC06459.1; -; Genomic_DNA.
DR   PIR; C70309; C70309.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR012287; Homeodomain-rel.
DR   InterPro; IPR002197; HTH_Fis.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF02954; HTH_8; 1.
DR   Pfam; PF00158; Sigma54_activat; 1.
DR   PRINTS; PR01590; HTHFIS.
DR   TIGRFAMs; TIGR01199; HTH_fis; 1.
DR   PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
DR   PROSITE; PS50045; SIGMA54_INTERACT_4; 1.
```

```
KW   Complete proteome.
SQ   SEQUENCE   210 AA;  23948 MW;  29E0F23B11FD1A77 CRC64;

  Query Match             16.4%;  Score 91;  DB 2;  Length 210;
  Best Local Similarity   28.4%;  Pred. No. 3.3;
  Matches   40;  Conservative   18;  Mismatches   37;  Indels   46;  Gaps    7;

Qy          1 RKAD------TKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSS----- 49
              ||||      | |||::  | |    |||| || :  :|||   ||   |||  |
Db         41 RKADVRIIAATNKNLKKLVEEGK-FREDLYYRLNVVNVELPPLRERKEDIPHLVSYFLEK 99

Qy         50 --------------------LPQDNRGNSRDSKEISIIEKTNRESI---TTNVEGRRDIH 86
                               |  |  || |:      :|     ::   :|:: |  |:
Db        100 FSKIHGKKVKGVSPEAMKLLLEYDYPGNVRE------LENAIEHAVVVSSTSLIGPDDLP 153

Qy         87 KGHLEEKKDGS-IKPEQKEDK 106
                   || : || :| : :|::
Db        154 ----EEIRIGSKVKKKFRENE 170


RESULT 25
Q7RQI4_PLAYO
ID   Q7RQI4_PLAYO PRELIMINARY;      PRT;   354 AA.
AC   Q7RQI4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=PY01113;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01000295; EAA20366.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   354 AA;  41675 MW;  700A43802C5EA7F1 CRC64;

  Query Match             15.6%;  Score 86.5;  DB 2;  Length 354;
  Best Local Similarity   28.2%;  Pred. No. 15;
  Matches   33;  Conservative   18;  Mismatches   37;  Indels   29;  Gaps    6;

Qy          7 KNLERKKEHGDILAEDLYGRLEI-PAIELPSENERGYYIPHQSSLPQDNRGNSRDSKE-- 63
              | |:: ||  |||        : |  | |:  : |:   |     :   :|  || | |:
Db         11 KLLKKNKERMDIL-------MGISKATEIEKDEEKKLKIEKNENADDNNNQNSNDKKQEN 63

Qy         64 -------ISIIEKTNRESIT-TNVEGRRDI-----HKGHLEEKKDGSIKPEQKEDKS 107
                    ::|: :||   :| :| :   ||      |    |:| ||      ||||:
Db         64 EKLKNSNLNIMNETNTSPVTVSNSDTINDIDLNANHXNGKEDKTDG------KEDKT 114


RESULT 26
Q7RJN5_PLAYO
ID   Q7RJN5_PLAYO PRELIMINARY;      PRT;   710 AA.
AC   Q7RJN5;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Cell cycle control protein cwf2.
GN   Name=PY03224;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR   EMBL; AABL01000919; EAA22781.1; -; Genomic_DNA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0000398; P:nuclear mRNA splicing, via spliceosome; IEA.
DR   InterPro; IPR012677; a_b_plait_nuc_bd.
DR   InterPro; IPR000504; RNP1_RNA_bd.
DR   Pfam; PF00076; RRM_1; 1.
DR   PROSITE; PS50102; RRM; 1.
KW   mRNA processing; mRNA splicing.
SQ   SEQUENCE   710 AA;  80642 MW;  90AB6EBAB1851F1E CRC64;

  Query Match             15.3%;  Score 85;  DB 2;  Length 710;
  Best Local Similarity   28.7%;  Pred. No. 48;
  Matches   37;  Conservative   17;  Mismatches   41;  Indels   34;  Gaps    7;

Qy          2 KADTKKNLERKKEHG---DILAEDLYGRLEIPAIELPSENERGYYIPHQSSL--PQDNRG 56
              | :  || : | | |   |  :||  |::  |       ||:|  :   :|   |: :
Db         67 KNEDDKNKDDKNEDGKNEDDKSED--GKVCEP-------NEKGEELKEGNSKENPKKRKK 117

Qy         57 NSRDSKEISIIEKTNRES-----ITTNVEGRRDIHKGHLEEKKDGSIKPE---------- 101
              ||   ||   :|| | |:         |:    |: :   | |||| | |
Db        118 NSETIKEDDNVEKDNNENGSKKKAKCNISIEEDLKES--ENKKDGSEKGEPVCNSKSERS 175

Qy        102 ---QKEDKS 107
                 :| :||
Db        176 EKGEKSEKS 184


RESULT 27
Q4RLC8_TETNG
ID   Q4RLC8_TETNG PRELIMINARY;      PRT;  3812 AA.
AC   Q4RLC8;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Chromosome 21 SCAF15022, whole genome shotgun sequence.
DE   (Fragment).
GN   ORFNames=GSTENG00032545001;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
```

```
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA   Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA   Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA   Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA   Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA   Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA   Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA   Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA   Kellis M., Volff JN., Guigo R., Zody M.C., Mesirov J.,
RA   Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA   Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA   Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT   "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT   the early vertebrate proto-karyotype.";
RL   Nature 431:946-957(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RG   Genoscope; Whitehead Institute Centre for Genome Research;
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; CAAE01015022; CAG10804.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   3812 AA;  435924 MW;  8D685142D3795352 CRC64;

  Query Match             15.3%;  Score 85;  DB 2;  Length 3812;
  Best Local Similarity   30.6%;  Pred. No. 3.5e+02;
  Matches   33;  Conservative   16;  Mismatches   39;  Indels   20;  Gaps    6;

Qy          2 KAD-TKKNLERKKEHGDILAEDLYGRLEIPAIEL-----------PSENERGYYIPHQS 48
              | | ||: |:  ::: : | |||: | ::   ||             :| |   |  ||
Db       2486 KVDSTKEELDANRQYSEKLQEDLHLR-DLSVSELQQELQKLRENLNTTEEEFSIYKEHQG 2544

Qy         49 SLPQDNRGNSRDSKEISIIEK-TNRESITTNVEGRRDIHKGHLEEKKD 95
              | :| :| |    |: : ||  | ||      |    ||| | |  ::
Db       2545 KLHEDLQG-SMSKLEVELQEKRQNLES----TEKELGIHKQHNENLQE 2587


RESULT 28
Q6R1M2_PLAFA
ID   Q6R1M2_PLAFA PRELIMINARY;      PRT;   429 AA.
AC   Q6R1M2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Merozoite surface protein 6 (Fragment).
GN   Name=MSP6;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=15039357; DOI=10.1128/IAI.72.4.2321-2328.2004;
RA   Pearce J.A., Triglia T., Hodder A.N., Jackson D.C., Cowman A.F.,
RA   Anders R.F.;
RT   "Plasmodium falciparum merozoite surface protein 6 is a dimorphic
RT   antigen.";
RL   Infect. Immun. 72:2321-2328(2004).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   Pearce A.J., Triglia T., Hodder A.N., Jackson D.C., Cowman A.F.,
RA   Anders R.F.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY518891; AAS13449.1; -; Genomic_DNA.
DR   InterPro; IPR010784; Merozoite_SPAM.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF07133; Merozoite_SPAM; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW   Merozoite.
FT   NON_TER     429    429
SQ   SEQUENCE   429 AA;  48666 MW;  9AB476FBF1CA078D CRC64;

  Query Match             15.3%;  Score 84.5;  DB 2;  Length 429;
  Best Local Similarity   29.0%;  Pred. No. 29;
  Matches   27;  Conservative   17;  Mismatches   44;  Indels    5;  Gaps    2;

Qy         17 DILAEDLYGRLEIPAIE---LPSENERGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRE 73
              |   | |  :::||: :   :| |||:    | : :  :|   :      |    | | |
Db        255 DKKTEYLLEQIKIPSWDRNNIPDENEQVKEDPQEDNKDEDEDEDEETETENLETEDDNNE 314

Qy         74 SITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDK 106
              |  | |   || :  :|||::   | |::| |
Db        315 EIEENEED--DIDEESVEEKEEEEEKKEEEEKK 345


RESULT 29
Q8R5C4_MOUSE
ID   Q8R5C4_MOUSE PRELIMINARY;      PRT;   457 AA.
AC   Q8R5C4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Expressed sequence C80913.
GN   Name=C80913;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N-3;
RC   TISSUE=Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by
RC   biopsy.;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=FVB/N-3;
RC   TISSUE=Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by
RC   biopsy.;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC023029; AAH23029.1; -; mRNA.
DR   MGI; MGI:1342294; C80913.
DR   GO; GO:0016272; C:prefoldin complex; IEA.
DR   GO; GO:0051082; F:unfolded protein binding; IEA.
DR   GO; GO:0006457; P:protein folding; IEA.
DR   InterPro; IPR004127; PFD_alpha_like.
DR   Pfam; PF02996; Prefoldin; 1.
SQ   SEQUENCE   457 AA;  50601 MW;  A8C7C719EE881902 CRC64;

  Query Match             15.2%;  Score 84;  DB 2;  Length 457;
  Best Local Similarity   26.1%;  Pred. No. 35;
```

```
  Matches   30;  Conservative   21;  Mismatches   40;  Indels   24;  Gaps    4;

Qy          3 ADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENER-------GYYIPHQSSLPQDNR 55
              :: |:  :||:: |       :   |:|||: |::  |       | |:||:| |   :|
Db        283 SEKKEEAKRKRKSG----AGSHATHELPAIKSPADIYRVFVDVVNGEYVPHKSILKSRSR 338

Qy         56 GNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEE-----KKDGSIKPEQKED 105
               ||        :     ||   :|| || : :    |     :  ||   | :|:
Db        339 ENS--------VCSDTSESSAADVEDRRGLLRSTSSEEAVATEAGGSSLDELQEN 385


RESULT 30
Q56PC1_CALJA
ID   Q56PC1_CALJA PRELIMINARY;      PRT;   596 AA.
AC   Q56PC1;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   RPGR (Fragment).
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini;
OC   Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Shu X., Fry A.M., Tulloch B., Manson F.D., Crabb J.W., Khanna H.,
RA   Faragher A., Lennon A., He S., Trojan P., Giessl A., Wolfrum U.,
RA   Swaroop A., Vervoort R., Wright A.F.;
RT   "RPGR ORF15 isoform co-localizes with RPGRIP1 at centrioles and basal
RT   bodies and interacts with nucleophosmin.";
RL   Submitted (DEC-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY855164; AAX49606.1; -; mRNA.
FT   NON_TER       1      1
SQ   SEQUENCE   596 AA;  66115 MW;  E81FF37BBF7A3448 CRC64;

  Query Match             15.2%;  Score 84;  DB 2;  Length 596;
  Best Local Similarity   27.4%;  Pred. No. 48;
  Matches   34;  Conservative   18;  Mismatches   36;  Indels   36;  Gaps    6;

Qy          4 DTKKNLE------RKKEHGDILAEDLYGRLEIPAIELPS--ENERGYYIPHQSSLPQDNR 55
              : | | |      |:||  :||::||  ::|:   :  |  | | |          : |
Db         20 EVKANEENVKVHGRRKEETEILSDDLTDKVEVSEGKAKSGGEAEDG----------PEGR 69

Qy         56 GNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLE--------------EKKDGSIKP 100
              |:    :| |  |    |  |   :| :|  :| :|              ||:||  :
Db         70 GDGTSEEESSATEHWQGEERTE--KGEKDKGRGEIEKPGEGEKDLPDKEEWEKRDGE-EQ 126

Qy        101 EQKE 104
              ||||
Db        127 EQKE 130


Search completed: December 15, 2005, 21:48:43
Job time : 232 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 250.224 Seconds
(without alignments)
5204.614 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22895/app_query.fasta_1.3740
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_821_@runat_21122005_093215_22895 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2478	99.7	493	2	AAR26944	Aar26944 P.falcipa
2	1640.5	66.0	456	8	ADO21942	Ado21942 LSA-NRC(H
3	1630	65.6	457	8	ADO21920	Ado21920 LSA-NRC(H
4	1561	62.8	316	2	AAR26941	Aar26941 P.falcipa
5	969.5	39.0	462	2	AAR05766	Aar05766 Portion o
6	770.5	31.0	318	2	AAR26943	Aar26943 P.falcipa
7	603	24.3	117	2	AAR26937	Aar26937 P.falcipa
c 8	376.5	14.3	419	4	ABG17301	Abg17301 Novel hum
9	376	15.1	1898	2	AAY30795	Aay30795 A human t
10	376	15.1	1898	7	ADD48869	Add48869 Human Pro
11	374	15.0	1439	8	ADR12189	Adr12189 Trichohya
12	374	15.0	1439	8	ADU77050	Adu77050 Murine ha
13	374	15.0	1439	9	ADW00404	Adw00404 Epithelia
14	373	15.0	1584	8	ABM80133	Abm80133 Tumour-as
15	361.5	14.5	412	2	AAW03626	Aaw03626 Human thy
16	359	14.4	1162	3	AAY96255	Aay96255 Kaposi's
17	359	14.4	1162	3	AAY58500	Aay58500 HHV8 ORF
18	359	14.4	1162	4	AAB62331	Aab62331 Amino aci
19	359	14.4	1162	5	ABB05621	Abb05621 Kaposi's
20	359	14.4	1162	8	ADJ65096	Adj65096 HHV8 late
21	359	14.4	1162	9	ADV68154	Adv68154 Kaposi's
22	339.5	13.7	207	3	AAB44672	Aab44672 Gene 33 h
23	335	13.5	562	2	AAR70491	Aar70491 Leucocyto
24	332.5	13.4	1192	3	AAB18165	Aab18165 Plasmodiu
25	331	13.3	2274	4	ABB58657	Abb58657 Drosophil
26	330.5	13.3	2990	8	ADP29884	Adp29884 Human sec
27	327.5	13.2	360	2	AAW03627	Aaw03627 Human fol
28	323.5	13.0	611	2	AAY29039	Aay29039 T. gondii
29	323.5	13.0	611	4	AAU25510	Aau25510 T. gondii
30	323.5	13.0	611	7	ADG17131	Adg17131 T. gondii
31	323	13.0	345	4	ABU53267	Abu53267 Human tes
32	320	12.9	658	4	ABB65632	Abb65632 Drosophil
33	320	12.9	800	4	ABB71459	Abb71459 Drosophil
34	320	12.9	3616	8	ADF45530	Adf45530 Chicken A
35	319	12.8	1972	2	AAW00024	Aaw00024 Smooth mu
36	318	12.8	554	4	ABB59454	Abb59454 Drosophil
37	317.5	12.8	1937	8	ADQ17289	Adq17289 Human sof
38	317.5	12.8	1937	8	ADQ17242	Adq17242 Human sof
39	316	12.7	1180	7	ADF06151	Adf06151 Bacterial
40	315.5	12.7	1170	5	ABB05710	Abb05710 Human tes
41	315.5	12.7	1170	7	ADC06789	Adc06789 Human pro
42	315.5	12.7	1170	7	ADC06845	Adc06845 Prostate
43	315.5	12.7	1501	5	ABP69798	Abp69798 Human pol
44	315.5	12.7	1564	6	ABP98812	Abp98812 Human str
45	315.5	12.7	1564	7	ADC06788	Adc06788 Human pro

ALIGNMENTS

RESULT 1
AAR26944
ID AAR26944 standard; protein; 493 AA.
XX
AC AAR26944;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA gene C-terminal region.
XX
KW Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW paludism; liver stage-specific antigen.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT Region 13. .213
FT /label= repetitive_region
FT Region 214. .493
FT /label= non-repetitive_region
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.
XX
PR 05-FEB-1991; 91FR-00001286.
XX
PA (INSP) INST PASTEUR.
XX
PI Guerinmarchand C, Druilhe P;

XX
DR WPI; 1992-299985/36.
DR N-PSDB; AAQ28119.
XX
PT Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT vaccination against, treatment of and diagnosis of malaria.
XX
PS Claim 2; Fig 8-10; 81pp; French.
XX
CC The 3' part of the P.falciparum liver-stage specific antigen (LSA) gene
CC codes for a polypeptide sequence which carries a T cell epitope
CC characteristic of a protein produced in hepatocytes infected with
CC P.falciparum. The polypeptide can be used in the preparation of vaccines
CC against malaria. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 493 AA;

Alignment Scores:

Pred. No.:	2.81e-177	Length:	493
Score:	2478.00	Matches:	493
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.68%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-41 (1-1482) x AAR26944 (1-493)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 20

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 40

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 60

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 80

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGlu 100

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGln 120

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 140

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLys 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        221 AsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu 240

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 260

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 280

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 300

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 320

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGln 340

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGlu 360

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAsp 380

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeu 400

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLys 420

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeu 440

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 TyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLys 460

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 PheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAsp 480

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||
Db        481 GluLeuSerGluAspIleThrLysTyrPheMetLysLeu 493
```

RESULT 2
ADO21942
ID ADO21942 standard; protein; 456 AA.
XX
AC ADO21942;
XX
DT 12-AUG-2004 (first entry)
XX
DE LSA-NRC(H) construct protein derived from Malaria parasite LSA-1.
XX
KW LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW malaria parasite P. falciparum; LSA-NRC(H) construct.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO2004044167-A2.
XX

PD 27-MAY-2004.
XX
PF 12-NOV-2003; 2003WO-US036011.
XX
PR 12-NOV-2002; 2002US-0425719P.
XX
PA (REED-) REED ARMY INST RES WALTER.
XX
PI Lanar DE, Hillier CJ, Lyon JA, Angov E, Kumar S, Rogers W;
PI Barbosa A;
XX
DR WPI; 2004-420309/39.
DR N-PSDB; ADO21941.
XX
PT Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT (LSA-1) epitope.
XX
PS Claim 4; SEQ ID NO 26; 90pp; English.
XX
CC The invention relates to a novel recombinant LSA-NRC polypeptide
CC comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC falciparum. The polypeptide of the invention demonstrates antimalarial
CC activities and may be useful as a vaccine during immunotherapy and as a
CC diagnostic agent, preferably for diagnosing malaria. The current sequence
CC is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC construct is harmonised for expression in Escherichia coli and comprises
CC the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC parasite P. falciparum LSA-1 in addition to a C-terminal His6 tag.
XX
SQ Sequence 456 AA;

Alignment Scores:

Pred. No.:	1.56e-114	Length:	456
Score:	1640.50	Matches:	338
Percent Similarity:	79.65%	Conservative:	26
Best Local Similarity:	73.96%	Mismatches:	70
Query Match:	65.99%	Indels:	23
DB:	8	Gaps:	5

US-09-837-344-41 (1-1482) x ADO21942 (1-456)

```
Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||::::::   :::  |||     |||:::     |||:::
Db          6 GluLysAspGluIleIleLysSerAsnLeuArgSerGlySerSerAsn------SerArg 23

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
                 |||      |||:::      |||::::::      |||
Db         24 AsnArgIleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerTyrGlu 43

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---------------- 288
              |||   |||      |||              ::::::|||:::
Db         44 LysThrLysAsnAsnGluAsnAsnLysPhePheAspLysAspLysGluLeuThrMetSer 63

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                 |||      |||              |||   :::
Db         64 AsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer 83

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
              |||   :::      :::::::::   |||   :::|||   :::      |||   :::
Db         84 GluAsnIlePheLeuLysGluAsnLysLeuAsnLysGluGlyLysLeuIleGluHisIle 103

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
                      |||      :::::::::      |||   :::   :::   :::
Db        104 IleAsnAspAspAspAsp------LysLysLysTyrIleLysGlyGlnAspGluAsnArg 121

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
              |||   |||||||||                |||||      ||||||||||||||||
Db        122 GlnGluAspLeuGlu----------------GluLysAlaAlaGluGlnGlnSerAsp 135

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
              ||||||||||||||   ||||||||||||||                          |||
```

```
Db        136 LeuGluGlnGluArgLeuAlaLysGluLysLeu----------------------Gln 147

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAla 167

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db        168 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAlaGluAspLeu 187

Qy        709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 207

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 IleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGlu 227

Qy        829 ATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg 247

Qy        889 GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAA 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 AspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLys 267

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 GluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspVal 287

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 AsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeu 307

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 IleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAsp 327

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLys 347

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGlu 367

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsn 387

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLys 407

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeu 427

Qy       1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 444
```

RESULT 3
ADO21920
ID ADO21920 standard; protein; 457 AA.
XX
AC ADO21920;
XX
DT 12-AUG-2004 (first entry)
XX
DE LSA-NRC(H)Mut construct protein derived from Malaria parasite LSA-1.
XX
KW LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;

```
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H)Mut construct; mutant; mutein.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 433
FT                   /note= "Residue inserted relative to wild-type P.
FT                   falciparum T5 epitope"
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   N-PSDB; ADO21919.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 5; SEQ ID NO 4; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to an amino acid insertion
CC   within the T5 P. falciparum epitope and a C-terminal His6 tag.
XX
SQ   Sequence 457 AA;

Alignment Scores:
Pred. No.:              9.56e-114      Length:         457
Score:                  1630.00        Matches:        338
Percent Similarity:     79.48%         Conservative:   26
Best Local Similarity:  73.80%         Mismatches:     70
Query Match:            65.57%         Indels:         24
DB:                     8              Gaps:           6

US-09-837-344-41 (1-1482) x ADO21920 (1-457)

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||::::::   :::   |||      |||:::         |||:::
Db          6 GluLysAspGluIleIleLysSerAsnLeuArgSerGlySerSerAsn------SerArg 23

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
                  |||      |||:::      |||::::::      |||
Db         24 AsnArgIleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerTyrGlu 43

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA--------------- 288
              |||   |||      |||                 ::::::|||:::
Db         44 LysThrLysAsnAsnGluAsnAsnLysPhePheAspLysAspLysGluLeuThrMetSer 63

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                  |||         |||                |||   :::
Db         64 AsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer 83

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
              |||   :::      :::::::::   |||   :::|||   :::      |||   :::
Db         84 GluAsnIlePheLeuLysGluAsnLysLeuAsnLysGluGlyLysLeuIleGluHisIle 103

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
                        |||      :::::::::      |||   :::   :::   :::
Db        104 IleAsnAspAspAspAsp------LysLysLysTyrIleLysGlyGlnAspGluAsnArg 121

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
              |||   |||||||||                  ||||||      ||||||||||||||
Db        122 GlnGluAspLeuGlu-----------------GluLysAlaAlaGluGlnGlnSerAsp 135

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
              |||||||||||||||   |||||||||||||||                        |||
Db        136 LeuGluGlnGluArgLeuAlaLysGluLysLeu-----------------------Gln 147

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAla 167

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
              |||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        168 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAlaGluAspLeu 187

Qy        709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 207

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 IleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGlu 227

Qy        829 ATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg 247

Qy        889 GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAA 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 AspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLys 267

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 GluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspVal 287

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 AsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeu 307

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 IleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAsp 327

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLys 347

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGlu 367

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsn 387

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLys 407

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeu 427

Qy       1429 CAGATCGTGGATGAG---TTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
```

```
                ||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        428 GlnIleValAspGluArgLeuSerGluAspIleThrLysTyrPheMetLysLeu 445

RESULT 4
AAR26941
ID   AAR26941 standard; protein; 316 AA.
XX
AC   AAR26941;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA-R-NR protein.
XX
KW   Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism; liver stage-specific antigen.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .209
FT                   /label= repeat_region
FT                   /note= "contains 12 x 17mer repeats"
FT   Region          210. .316
FT                   /label= non-repeat_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28115.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Disclosure; Fig 1; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert (950 bases) encoded LSA-R-NR
CC   containing a 12-repeat region followed by a non-repeat region. Preferred
CC   antigenic polypeptides of the invention are derived from the amino acid
CC   sequence of LSA-R-NR. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 316 AA;

Alignment Scores:
Pred. No.:              1.35e-108      Length:         316
Score:                  1561.00        Matches:        316
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            62.79%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x AAR26941 (1-316)

Qy          13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy          73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy         133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy         193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy         253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy         313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy         373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy         433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         141 GlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy         493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy         553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy         613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAATTTAGAAAGA 672
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy         673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy         733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy         793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy         853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy         913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 5
AAR05766
ID   AAR05766 standard; protein; 462 AA.
XX
AC   AAR05766;
XX
DT   25-MAR-2003  (revised)
DT   05-NOV-1990  (first entry)
XX
DE   Portion of peptide antigen to malarial sporozite.
XX
KW   Malaria; sporozite; vaccine; exoerythrocytic parasites; tetanus toxoid.
XX
```

```
OS   Plasmodium falciparum.
XX
PN   WO9006130-A.
XX
PD   14-JUN-1990.
XX
PF   30-NOV-1988;   88US-00278234.
XX
PR   30-NOV-1988;   88US-00278234.
PR   12-APR-1989;   89US-00337204.
XX
PA   (BIOM-) BIOMEDICAL RES INST.
XX
PI   Hollingdal MR;
XX
DR   WPI; 1990-209624/27.
DR   N-PSDB; AAQ05140.
XX
PT   Novel malarial sporozoite antigenic protein - useful as vaccine against
PT   sporozoite(s) and exo:erythrocytic parasites.
XX
PS   Disclosure; Page ?; -pp; English.
XX
CC   Antigen, preferably linked at the C-terminal to a carrier such as tetanus
CC   toxoid, may be used as a vaccine against the malarial sporozite. (Updated
CC   on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 462 AA;

Alignment Scores:
Pred. No.:              2.94e-64        Length:         462
Score:                  969.50          Matches:        232
Percent Similarity:     61.88%          Conservative:   57
Best Local Similarity:  49.68%          Mismatches:     95
Query Match:            39.00%          Indels:         83
DB:                     2               Gaps:           10

US-09-837-344-41 (1-1482) x AAR05766 (1-462)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||   ||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db         74 GlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 93

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db         94 GlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 113

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 LeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 133

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||:::   :::
Db        134 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluArgThrLysAla 153

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              :::   |||   |||   |||||||||||||||||||||||:::|||||||||||||||
Db        154 SerThrGluThrLeuHisGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGlu 173

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              |||||||||   |||||||||||||||||||||||||   ||||||||||||||||||
Db        174 LysLeuGlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 193

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        194 GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGln 213

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        214 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 233

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||:::|||   ||||||||||||||||||||||||||||||||||||||||||:::

Db        234 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAsp 253

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||   |||
Db        254 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAla 273

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||   ||||||   |||||||||:::   ||||||   |||:::
Db        274 LysGluLysLeuGlnValGlnGlnSerAspLeuGluArgThrLysAlaSerThrGluThr 293

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                  ||||||
Db        294 LeuGlnGluArg---------------------------------------------- 297

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                                                                    |||
Db        298 ----------------------------------------------------------Gln 298

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||   |||   |||:::   |||         :::     ::::::   |||   :::
Db        299 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 318

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::       |||                |||   :::||||||:::   :::
Db        319 GluGlnGluArgArgAlaLysGluLysLeuGln---GluGlnGlnArgAspLeuGluGln 337

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                  |||   :::                        |||:::   :::::::::
Db        338 GluArgLeuAlaLys----------------------GluLysLeuGlnGluGlnGln 349

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              :::|||::::::         |||         :::
Db        350 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeu---------------------- 362

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                                                                    :::
Db        363 ----------------------------------------------------------Gln 363

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||         ||||||::::::   :::      :::      :::      |||:::
Db        364 GluGlnGlnArgAspLeuGluGlnGluArg------LeuAlaLysGluLysLeuGlnGlu 381

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::::::::              :::||||||                  |||   |||
Db        382 GlnGlnSer--------------AspLeuGluGlnGluArgArgAlaLysGluLysLeu 396

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
                 :::            :::|||:::      |||   |||::::::   |||::::::
Db        397 GlnGluGlnGlnSerAspLeuGluGln------GluArgLeuAlaLysValLysLeuGln 414

Qy       1261 AAAGGAAAGAAATAT---GAAAAAACAAAG------GATAATAATTTTAAACCAAATGAT 1311
              :::   :::   |||   |||:::||||||                  :::
Db        415 GluGlnGlnSerTyrLeuGluArgThrLysAlaSerThrGluThrLeuGlnGluArgGln 434

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAA---AATGATAAGCAGGTTAATAAG 1368
                    |||       |||   :::   |||   |||      ::::::|||   :::
Db        435 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 454

Qy       1369 GAAAAGGAAAAATTCATAAAA 1389
              |||:::|||:::      |||
Db        455 GluGlnGluArgArgAlaLys 461

RESULT 6
AAR26943
ID   AAR26943 standard; protein; 318 AA.
XX
AC   AAR26943;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA N-terminal portion.
XX
```

```
KW   Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW   paludism; liver stage-specific antigen; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .153
FT                   /label= unique_5'_region
FT   Region          154. .318
FT                   /label= repetitive_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28117.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 11; Fig 7; 81pp; French.
XX
CC   The 5' part of the P.falciparum liver-stage specific antigen (LSA) gene
CC   codes for a polypeptide sequence which carries a T cell epitope
CC   characteristic of a protein produced in hepatocytes infected with
CC   P.falciparum. The polypeptide can be used in the preparation of vaccines
CC   against malaria. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 318 AA;

Alignment Scores:
Pred. No.:              2.28e-49      Length:          318
Score:                  770.50        Matches:         163
Percent Similarity:     92.59%        Conservative:    12
Best Local Similarity:  86.24%        Mismatches:      13
Query Match:            30.99%        Indels:          1
DB:                     2             Gaps:            1

US-09-837-344-41 (1-1482) x AAR26943 (1-318)

Qy          16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 75
               |||   ::::::::::::      |||   :::   :::   :::|||   |||||||||
Db         131 AspAspAspLysLysLysTyrIleLysGlyGlnAspGluAsnArgGlnGluAspLeuGlu 150

Qy          76 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 135
                  ::::::   ||||||||||||||||||   ||||||||||||   ||||||||||||
Db         151 ---GluLysAlaAlaLysGluLysLeuGlnGlyGlnGlnSerAspSerGluGlnGluArg 169

Qy         136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         170 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLys 189

Qy         196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         190 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 209

Qy         256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
               ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db         210 GlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 229

Qy         316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         230 GlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAsp 249

Qy         376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         250 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 269

Qy         436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         270 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeu 289

Qy         496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         290 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 309

Qy         556 AAGTTGCAAGAACAACAAAGCGATTTA 582
               :::|||||||||||||||||||||||||
Db         310 ArgLeuGlnGluGlnGlnSerAspLeu 318

RESULT 7
AAR26937
ID   AAR26937 standard; peptide; 117 AA.
XX
AC   AAR26937;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum liver-stage specific antigen non-repeat region.
XX
KW   Malaria; LSA-NR; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism.
XX
OS   Plasmodium falciparum.
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 4; Page 55; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert encoded LSA-R-NR containing a
CC   12-repeat region (LSA-R) followed by a non-repeat region (LSA-NR). The
CC   sequence given here is a preferred polypeptide of the invention carrying
CC   a T cell epitope typical of liver-stage P.falciparum. See AAR26919-
CC   R26935, AAR26937 and AAR26939. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 117 AA;

Alignment Scores:
Pred. No.:              7.16e-37      Length:          117
Score:                  603.00        Matches:         117
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            24.26%        Indels:          0
DB:                     2             Gaps:            0
```

```
US-09-837-344-41 (1-1482) x AAR26937 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 8
ABG17301
ID   ABG17301 standard; protein; 419 AA.
XX
AC   ABG17301;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #17292.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS81488.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 47660; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 419 AA;

Alignment Scores:
Pred. No.:              7.81e-20        Length:         419
Score:                  376.50          Matches:        149
Percent Similarity:     44.34%          Conservative:   35
Best Local Similarity:  35.90%          Mismatches:     76
Query Match:            14.26%          Indels:         155
DB:                     4               Gaps:           25

US-09-837-344-41 (1-1482) x ABG17301 (1-419)

Qy       1217 CCTTCATCTAAATCATCTAAATTTTCATTTTTCTCTATCAAATCTTCTAGTTCTTTATAA 1158
              |||:::      ||||||      |||      |||:::   ||||||   ||||||
Db        136 ProAlaProAlaSerSerProCysSer---ArgSerLeuSerSerSerArgSerLeu--- 153

Qy       1157 ATTCCTATGTTTTCTTCATCTTGGAAA-----------------TTGTCATATTGCACA 1116
                          |||||||||   :::                   ||||||   |||
Db        154 -----------SerSerSerCysArgValSerCysCysHisSerLeuSerSerCysCys 169

Qy       1115 ATAGGCTTAAATTCGTCTAAGTCTTCATCATCTTCTTCTTCATCTATTAATGAATCGTCA 1056
                    |||:::   |||:::|||||||||:::||||||   |||   :::   ||||||
Db        170 ArgSerLeuSerCysSerArgSerSerSerThrSerSerArgSerSerSerThrSerSer 189

Qy       1055 TATTCAGCACTTATTTCATCCTCATACTTACTTATTTGAAAATCATTAACATCAGAAATA 996
                 |||      :::|||||||||
Db        190 SerSerLeuSerLeuSerSerSerLeu---------------------------------- 198

Qy        995 TTTACTGTCTCTAATGTATGATTTTGTATGTCAGCAGATTTATCTTCTTTTTGTTCTGGT 936
                                        |||   |||      ||||||      ||||||
Db        199 --------------------SerCysSerSerCysArgLeuSer------CysSer--- 208

Qy        935 TTTATTGAACCATCTTTCTTTTCTTCAAGATGTCCTTTATGTATATCCCTTCGTCCTTCA 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 ---------------------------------------------------------- 208

Qy        875 ACATTTGTTGTAATAGATTCTCTATTTGTTTTTTCTATTATAGATATTTCCTTGGAATCT 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 ---------------------------------------------------------- 208

Qy        815 CTACTATTCCCTCTGTTGTCCTGAGGTAAAGAAGATTGATGTGGTATATAATATCCACGT 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 ----------------------------------------------------------Asn 209

Qy        755 TCATTTTCTGATGGAAGTTCTATAGCTGGTATTTCTAAACGACCATATAAATCCTCTGCT 696
              |||||||||      ||||||        :::|||
Db        210 SerPheSerThrSerSerSerSerLeuSerLeuSer----------------------- 221

Qy        695 AATATATCTCCATGTTCCTTTTTTCTTTCTAAATTTTTTTTCGTATCAGCCTTCCTTTGT 636
                                                          |||:::      |||
Db        222 ----------------------------------------------SerSerLeuSerCys 226

Qy        635 TCTAAATCTCTTTGCTGCTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCG 576
              |||   |||   |||   ||||||:::   |||   :::      |||   |||   |||
Db        227 Ser---SerCysCysLeuSerCysSerAsnSerPheSerThrSerSerSerSerLeuSer 245

Qy        575 CTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGT 516
              |||      |||   :::   |||      ||||||||||||||   |||
Db        246 LeuSerSerSerLeuSerCysSerSerCysArgLeuSerCysSerAsnSerPhe------ 263
```

```
Qy        515 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAAC 456
                  :::   |||   :::   |||||||||      |||      ||||||   :::
Db        264 ------SerThrSerSerSerSerLeuSerCysProGlySerPheSerCysSerHisSer 281

Qy        455 TTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTA 396
                |||            |||||||||:::||||||      ||||||:::   ||||||
Db        282 LeuSer-----------SerCysSerArgSerLeu---SerSerCysSerHisSerLeu 296

Qy        395 GCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTC 336
                :::      |||||||||:::||||||      ||||||:::   ||||||:::
Db        297 Ser------SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------ 311

Qy        335 TCTTGTTCTAAATCGCTTTGCTGCTCTTGTAACTTTTCTTTAGCAAGTCTATCTTGTTCT 276
                |||||||||:::||||||      ||||||:::   ||||||:::      |||||||||
Db        312 SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSer 328

Qy        275 AAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTT 216
                :::||||||      ||||||:::   ||||||:::      |||||||||:::||||||
Db        329 ArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSerArgSerLeu 345

Qy        215 ---TGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTTTGTTGT 159
                   |||||||||:::   |||            ||||||||||||||||||
Db        346 SerSerCysSerCysSerLeuSer-----------SerCysSerArgSerLeu---Ser 360

Qy        158 TCTTGCAACTTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGCTGCTCTTGTAAC 99
                ||||||:::   ||||||:::      |||||||||   ||||||      ||||||:::
Db        361 SerCysSerHisSerLeuSer------SerCysSerHisSerLeu---SerSerCysSer 377

Qy         98 TTTTCTTTAGCAAGTCTATCTTGTTCTAAATCGCTTTGT-----------------TGT 57
                ||||||:::|||   |||||||||   |||   |||                   |||
Db        378 ArgSerLeuSerSerCysSerCysSerLeuSerSerCysSerArgCysLeuSerSerCys 397

Qy         56 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTT 12
                ||||||:::   |||            ||||||||||||||||||
Db        398 SerCysSerLeuSer-----------SerCysSerArgSerLeu 408
```

RESULT 9
AAY30795
ID AAY30795 standard; protein; 1898 AA.
XX
AC AAY30795;
XX
DT 25-NOV-1999 (first entry)
XX
DE A human trichohyalin (TRHY) protein.
XX
KW Human; trichohyalin; TRHY; protein; tissue structure; wound healing;
KW terminally differentiating epidermal tissue; proteinaceous gel;
KW breast implant.
XX
OS Homo sapiens.
XX
PN US5958752-A.
XX
PD 28-SEP-1999.
XX
PF 14-FEB-1997; 97US-00800644.
XX
PR 30-APR-1993; 93US-00056200.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Kim I, Chung S, Park S, Steinert PM, Lee S;
XX
DR WPI; 1999-561041/47.
DR N-PSDB; AAZ22301.
XX
PT Human trichohyalin useful for forming a proteinaceous gel that promotes
PT wound healing.
XX
PS Disclosure; Fig 3A-W; 126pp; English.
XX

CC The present sequence represents a human trichohyalin (TRHY) protein. The
CC protein is found in terminally differentiating epidermal tissue, and is
CC involved in forming the structural architecture of such tissue. The
CC trichohyalin protein is useful for forming a proteinaceous gel which may
CC then be used for healing wounds, or in breast implants
XX
SQ Sequence 1898 AA;

Alignment Scores:

Pred. No.:	1.02e-19	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	2	Gaps:	24

US-09-837-344-41 (1-1482) x AAY30795 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::          |||::::::||||||   :::              |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                ||||||   |||||||||   :::          :::|||::::::       ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
              ||||||||||||   :::          :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                 |||:::|||||||||   :::          :::||||||:::       ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::                  ||||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
              |||       |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                 ::::::||||||   :::|||:::   :::|||:::                :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 ----------------------------------GATTTAGAACAAGAG--------- 540
                                                 ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||       |||:::      |||:::                    :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
                ||||||::::::   |||||||||:::      |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793
```

```
Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
              |||   |||   |||            :::   |||      |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||   |||              :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::           |||      |||                  ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATATCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                       :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||      :::         |||:::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::            :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                  :::                 ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                |||          |||:::   |||||||||::::::         :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                  :::      :::   :::|||            |||:::
Db        935 ------------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||      |||   |||      |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||      :::     |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::           :::      :::|||              |||      ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 10
ADD48869
ID   ADD48869 standard; protein; 1898 AA.
XX
AC   ADD48869;
XX
DT   02-DEC-2004  (revised)
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein AAA65582, SEQ ID NO 14580.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
OS   Unidentified.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; AAA65582.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Example 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (described in Table 3
CC   of the specification) which is differentially expressed during pain.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1898 AA;

Alignment Scores:
Pred. No.:              1.02e-19        Length:         1898
Score:                  376.00          Matches:        137
Percent Similarity:     47.39%          Conservative:   117
Best Local Similarity:  25.56%          Mismatches:     174
Query Match:            15.12%          Indels:         108
DB:                     7               Gaps:           24

US-09-837-344-41 (1-1482) x ADD48869 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::||||||   :::            |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                  ||||||   |||||||||   :::         :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578
```

```
Qy      178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
            ||||||||||||   :::         :::|||||||:::
Db      579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy      220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                ||:::|||||||||   :::         :::||||||:::      ||||||
Db      599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy      280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
            |||   ||||||   :::|||:::::::::               ||||||||||||
Db      619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy      340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
            |||       |||:::|||                :::|||::::::   :::|||
Db      634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy      379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                 |||||||||   |||      |||:::|||:::   :::|||   :::
Db      654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy      430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA---------------CAAAGC 474
               ::::::||||||   :::|||:::   :::|||:::                :::
Db      674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy      475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
            :::   |||:::|||||||||   :::|||      :::|||:::|||
Db      694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy      526 -----------------------------------GATTTAGAACAAGAG--------- 540
                                                  ||||||   |||
Db      714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy      541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
            |||:::::::|||      |||:::      |||:::                     :::
Db      734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy      580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
               |||::::::||||||   :::   :::||||||
Db      754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy      616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
            ||||||::::::   |||||||||:::      |||   :::          |||
Db      774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy      673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
            |||   |||   |||             :::   |||       |||
Db      794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy      718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                    ::::::   |||   |||                  :::|||      :::
Db      811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy      769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
            :::|||   :::         |||       |||                  ||||||
Db      829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy      829 ATATCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
            :::      :::|||:::      :::                        :::   :::
Db      849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy      868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                ::::|||   :::   |||       :::          |||:::::::::|||
Db      869 GluGluGluAspGlyLeuGlnGluAspGlnArgArgArgArgGlnGluGlnArgArgAsp 888

Qy      925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                ::::   :::   :::|||::::::           :::   |||||||||
Db      889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy      985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                     :::               ::::::::::::
Db      907 ------------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy     1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                   |||           |||:::   |||||||||::::::         :::|||
Db      917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy     1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                     :::      :::   :::|||            |||:::
Db      935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy     1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
            |||   |||   |||   ::::::   |||   |||      |||   |||      |||
Db      949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy     1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
            |||         :::      |||   :::   :::|||   ||||||   ::::::
Db      967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy     1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
            ::::::          :::      :::|||            |||      ||||||
Db      987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy     1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
            |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db     1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022
```

RESULT 11
ADR12189
ID ADR12189 standard; protein; 1439 AA.
XX
AC ADR12189;
XX
DT 21-OCT-2004 (first entry)
XX
DE Trichohyalin related protein, SEQ ID 1.
XX
KW Hair growth; murine; trichohyalin.
XX
OS Mus musculus.
XX
PN WO2004063750-A1.
XX
PD 29-JUL-2004.
XX
PF 07-JAN-2004; 2004WO-JP000030.
XX
PR 08-JAN-2003; 2003JP-00001891.
XX
PA (SUME) SUMITOMO ELECTRIC IND LTD.
XX
PI Takebe K, Hirai Y;
XX
DR WPI; 2004-571498/55.
DR N-PSDB; ADR12190.
XX
PT Evaluating hair-promoting activity, by using antibody that recognizes
PT proteins (antigen) of epithelial newborn follicles.
XX
PS Claim 1; SEQ ID NO 1; 39pp; Japanese.
XX
CC The present invention relates to a method (M1) for evaluating hair growth
CC -promoting activity, by using an antibody that recognizes an antigen
CC (ADR12189) of epithelial newborn follicles. Also claimed is a kit for
CC evaluating hair growth-promoting activity, comprising the antibody of
CC (M1). The present sequence is a murine trichohyalin-related protein,
CC which was used as the antigen in the method of the invention.
XX
SQ Sequence 1439 AA;

Alignment Scores:

Pred. No.:	1.4e-19	Length:	1439
Score:	374.00	Matches:	143
Percent Similarity:	48.15%	Conservative:	130
Best Local Similarity:	25.22%	Mismatches:	170
Query Match:	15.04%	Indels:	124
DB:	8	Gaps:	25

```
US-09-837-344-41 (1-1482) x ADR12189 (1-1439)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
              :::|||::::::   :::|||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||      |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA---------------------------------------- 114
                   |||||:::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                  ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                |||||||||||      :::|||:::          |||||||||:::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||          |||   :::   ||||||::::::          |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              ::::::|||   :::::::::   ::::::|||:::          ||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                    |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                    :::|||:::|||:::::::::   |||      ||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                    |||   :::|||:::|||:::::::::          ||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                    |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::|||:::   :::||||||   ||||||         |||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   |||||||||:::::|||             |||   :::      :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
              :::            ||||||      :::   :::
Db        871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
              :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA----------------TCTATT 864
              :::|||:::   :::   ::::::         |||                  :::

Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                        |||   |||||||||       |||   |||   ||||||::::::
Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
                        :::   |||:::   |||:::|||         :::|||
Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ----------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                                          |||   |||             :::   |||
Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
                 :::|||   :::   |||   |||:::      |||   ::::::
Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
                                                     |||   :::|||:::
Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                 ::::::||||||   :::|||:::          :::   |||         :::|||
Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
              |||   :::          |||   :::|||   |||       |||
Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              :::   |||                :::::   :::   :::       |||
Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              |||:::::::::::::::   ::::::   |||:::   :::|||         :::
Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::              :::   |||   :::|||   :::::::::   :::   |||:::
Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377
              :::      :::|||:::|||
Db       1141 LysPheArgGluGluGlnGlu 1147

RESULT 12
ADU77050
ID   ADU77050 standard; protein; 1439 AA.
XX
AC   ADU77050;
XX
DT   10-FEB-2005  (first entry)
XX
DE   Murine hair growth stimulator related protein.
XX
KW   Hair growth stimulation; endocrine-gen.; cardiovascular-gen.;
KW   antidiabetic; hepatotropic; antiinflammatory; gastrointestinal-gen.;
KW   muscular-gen.; stimulator of hair growth; cardiovascular disease;
KW   diabetes; liver disease; periodontal disease; gastrointestinal disease;
KW   muscular dystrophy; cholecystitis.
XX
OS   Mus musculus.
XX
PN   WO2004101610-A1.
XX
PD   25-NOV-2004.
XX
PF   14-MAY-2004; 2004WO-JP006909.
XX
PR   16-MAY-2003; 2003JP-00138567.
PR   17-NOV-2003; 2003JP-00386185.
XX
```

PA (SUME) SUMITOMO ELECTRIC IND LTD.
XX
PI Hirai Y, Nakajima K;
XX
DR WPI; 2004-833974/82.
XX
PT Novel oligopeptide or its modified form useful for enhancing hair growth
PT or for treating cardiovascular diseases, diabetes, liver diseases,
PT periodontal disease, intestinal disease or muscular dystrophy.
XX
PS Disclosure; SEQ ID NO 11; 80pp; Japanese.
XX
CC The invention relates to an oligopeptide or its modified form, used for
CC promoting hair growth, comprising 5-100 amino acid residues. The
CC invention also relates to a composition comprising the oligopeptide, its
CC modified form or its salt, and a hair growth stimulant comprising the
CC oligopeptide, its modified form or its salt. The oligopeptide is useful
CC for enhancing hair growth or for treating cardiovascular diseases,
CC diabetes, liver diseases, periodontal disease, intestinal disease,
CC muscular dystrophy or cholecystitis. This sequence represents a murine
CC hair growth stimulation related protein of the invention.
XX
SQ Sequence 1439 AA;

Alignment Scores:

Pred. No.:	1.4e-19	Length:	1439
Score:	374.00	Matches:	143
Percent Similarity:	48.15%	Conservative:	130
Best Local Similarity:	25.22%	Mismatches:	170
Query Match:	15.04%	Indels:	124
DB:	8	Gaps:	25

US-09-837-344-41 (1-1482) x ADU77050 (1-1439)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------------ 48
              :::|||::::::   :::||||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||      |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                    ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                  ||||||::||||||       :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                |||||||||||      :::|||:::         |||||||||:::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
                |||   |||||         |||   :::   ||||||::::::         |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
                :::::|||   :::::::::   ::::::|||:::         ||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                     |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                     |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                     :::|||:::|||:::::::::   |||      ||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                     |||   :::|||:::|||:::::::::         ||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                     |||   :::|||:::|||:::::::::   |||      ||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                  :::|||:::   :::||||||   ||||||         |||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
                  :::   |||||||||:::::|||             |||   :::      :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                  :::            ||||||      :::   :::
Db        871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
                  :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
                  :::|||:::   :::   ::::::         |||                  :::
Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    |||   |||||||||      |||   |||   ||||||::::::
Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
                        :::   |||:::   |||:::|||         :::|||
Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ------------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                                       |||   |||         :::   |||
Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
                  :::|||   :::   |||   |||:::      |||   ::::::
Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ------------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
                                                   |||   :::|||:::
Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                  ::::::||||||   :::|||:::         :::   |||         :::|||
Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
                  |||   :::         |||   :::|||   |||         |||
Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                  :::   |||                ::::::   :::   :::      |||
Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
                  |||:::::::::::::::   ::::::   |||:::   :::|||         :::
Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                  :::            :::   |||   :::|||   :::::::::   :::   |||:::
Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140
```

```
Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377
              :::      :::|||:::|||
Db       1141 LysPheArgGluGluGlnGlu 1147

RESULT 13
ADW00404
ID   ADW00404 standard; protein; 1439 AA.
XX
AC   ADW00404;
XX
DT   07-APR-2005  (first entry)
XX
DE   Epithelial hair follicle related antigen, SEQ ID 157.
XX
KW   Endocrine-Gen.; Cardiovascular-Gen.; Vulnerary; Osteopathic; Cardiant;
KW   Gene Therapy; antibody production; antibody therapy; antigen;
KW   hair growth; arteriosclerosis; osteoporosis; bone injury; angina; burns;
KW   wound healing.
XX
OS   Mus musculus.
XX
PN   US2005008602-A1.
XX
PD   13-JAN-2005.
XX
PF   07-JAN-2004; 2004US-00754079.
XX
PR   05-JUN-2000; 2000JP-00166903.
PR   04-JUN-2001; 2001WO-JP004691.
PR   13-NOV-2001; 2001JP-00347338.
PR   13-NOV-2001; 2001JP-00347340.
PR   05-DEC-2001; 2001JP-00371175.
PR   05-DEC-2001; 2001JP-00371366.
PR   23-MAY-2002; 2002US-00155922.
PR   08-JAN-2003; 2003JP-00001891.
XX
PA   (SUME ) SUMITOMO ELECTRIC IND LTD.
XX
PI   Hirai Y,  Takebe K;
XX
DR   WPI; 2005-080487/09.
DR   N-PSDB; ADW00406.
XX
PT   New antibody specifically recognizing an antigen of 220kDa present in
PT   epithelial new follicles, useful for evaluating and inducing hair growth
PT   promoting activity and morphogenesis.
XX
PS   Claim 5; SEQ ID NO 157; 87pp; English.
XX
CC   The present invention relates to a novel antibody, or its fragment, which
CC   specifically recognizes an antigen of about 220kDa (ADW00404) present in
CC   new epithelial hair follicles. The antigen is specifically expressed
CC   during the growth period of an imago or the developing period of a fetus.
CC   Also claimed is a method for the evaluation of hair growth promoting
CC   activity, comprising incubating skin tissue derived from a mammal in the
CC   presence of a substance to be tested to promote hair growth, recovering
CC   the skin tissue, reacting the skin tissue with the antibody cited above,
CC   and detecting the antibody, or its fragment, that reacts with the skin
CC   tissue. The method and antibody of the invention are useful for
CC   evaluating and inducing hair growth promoting activity, treating and
CC   ameliorating symptoms of diseases associated with abnormal morphogenesis,
CC   inducing cardiovascular revascularization, regeneration and endothelial
CC   cell growth, such as in arteriosclerosis, Berger's disease, osteoporosis,
CC   fractures and angina pectoris, and in treating burns or wounds.
XX
SQ   Sequence 1439 AA;

Alignment Scores:
Pred. No.:              1.4e-19        Length:         1439
Score:                  374.00         Matches:        143
Percent Similarity:     48.15%         Conservative:   130
Best Local Similarity:  25.22%         Mismatches:     170
Query Match:            15.04%         Indels:         124

DB:                  9              Gaps:         25

US-09-837-344-41 (1-1482) x ADW00404 (1-1439)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
              :::|||::::::   :::|||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||         |||       |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                    ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                  ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
              ||||||||||||      :::|||:::         |||||||||::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||         |||   :::   ||||||::::::         |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              ::::::|||   :::::::::   ::::::|||:::         ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||       ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                    |||   :::|||:::|||:::::::::   |||       ||||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                    :::|||:::|||:::::::::   |||       ||||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                    |||   :::|||:::|||:::::::::         ||||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                    |||   :::|||:::|||:::::::::   |||       ||||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::|||:::   :::||||||   ||||||         |||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   |||||||||:::::|||             |||   :::       :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
              :::             ||||||      :::   :::
Db        871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
              :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
```

```
Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924

Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978

Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ----------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014

Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT---------- 1062

Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086

Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143

Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185

Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245

Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305

Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356

Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377

Db       1141 LysPheArgGluGluGlnGlu 1147
```

RESULT 14
ABM80133
ID ABM80133 standard; protein; 1584 AA.
XX
AC ABM80133;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO80641, SEQ:339.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR N-PSDB; ACN37449.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 12; SEQ ID NO 339; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 1584 AA;

Alignment Scores:

Pred. No.:	1.68e-19	Length:	1584
Score:	373.00	Matches:	133
Percent Similarity:	46.53%	Conservative:	115
Best Local Similarity:	24.95%	Mismatches:	185
Query Match:	15.00%	Indels:	100
DB:	8	Gaps:	21

US-09-837-344-41 (1-1482) x ABM80133 (1-1584)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db        278 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 297

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Db        298 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 314

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG---CAAGAACAACAAAGCGATCTAGAA 177

Db        315 ArgGluGlnGluGluArgArgGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 334

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------------------- 219

Db        335 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 354

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279

Db        355 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 374

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGCGATTTAGAACAA 333

Db        375 Gln---ArgLeuLysArgGluGluValGluArgLeuGluGlnGluGluArgArgGluGln 393
```

```
Qy       334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA-----------GAA 381
Db       394 ArgLeuLysArgGluGluProGluGluGluArgArgGlnGlnLeuLeuLysSerGluGlu 413

Qy       382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
Db       414 GlnGluGluArgArgGlnGlnGlnLeuArgArgGluGlnGlnGluArgArgGluGlnArg 433

Qy       442 CTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGCGATTTAGAACAA 486
Db       434 LeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArgGluHisGluGlu 453

Qy       487 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC-------------------- 525
Db       454 GluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArgGluArgIleLys 473

Qy       526 ----------------------GATTTAGAACAAGAG-------------------- 540
Db       474 SerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnSerLys 493

Qy       541 ---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591
Db       494 ValTyrSerArgProArgLysGlnGluGlyGlnArgArgArgGlnGluGlnGluLuLys 513

Qy       592 AGACGTGCTAAA--------------------------GAAAAGTTGCAAGAGCAGCAA 624
Db       514 ArgArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHisArgGlnGln 533

Qy       625 AGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGAAAAAAGGAA 681
Db       534 GlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGluLysSerGlu 553

Qy       682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGT----------------------- 717
Db       554 ArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGluGlnArgGlu 570

Qy       718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db       571 ArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------PheLeuProGlu 588

Qy       778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
Db       589 GluGluGluLysGluGlnArgArgArgGlnArgArgGluArgGluLysGluLeuGlnPhe 608

Qy       838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db       609 LeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGlnGluGluGlu 628

Qy       898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db       629 AspGly---LeuGlnGluAspGlnGluArgArgArgSerGlnGluGlnArgArgAspGln 647

Qy       958 TCTGCTGACATACAA----------------------AATCATACATTAGAGACAGTA 993
Db       648 LysTrpArgTrpGlnLeuGluGluGluArgLysArgArgArgHisThrLeu--------- 664

Qy       994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
Db       665 --------------------TyrAlaLysProAlaLeuGlnGluGlnLeuArgLysGlu 677

Qy      1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
Db       678 ------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu--------- 692

Qy      1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173
Db       693 ---------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArgGluGluGlu 709

Qy      1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
Db       710 GlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGluLysArgArg 727

Qy      1234 GAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293
```

```
Db       728 ArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLysGluGluGln 747

Qy      1294 AATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT 1350
Db       748 LeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLysTyrArgGlu 767

Qy      1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db       768 GluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 780
```

RESULT 15
AAW03626
ID AAW03626 standard; peptide; 412 AA.
XX
AC AAW03626;
XX
DT 25-MAR-2003 (revised)
DT 17-DEC-1996 (first entry)
XX
DE Human thyrotropin GPR N-terminal sequence.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
XX
OS Homo sapiens.
XX
PN US5508384-A.
XX
PD 16-APR-1996.
XX
PF 09-SEP-1993; 93US-00118270.
XX
PR 10-SEP-1992; 92US-00943236.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Schuster DI, Murphy RB;
XX
DR WPI; 1996-208785/21.
XX
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
PT treating schizophrenia.
XX
PS Disclosure; Fig 8B(2); 184pp; English.
XX
CC Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomegaloviral and other GPR proteins. The peptides AAW03578-
CC W03651 represent the N-terminal fragments of the above proteins. The
CC receptor proteins were used to design polypeptides, pref. based on the
CC transmembrane domains, for use in G-protein coupled receptor ligand
CC binding assays. The polypeptide fragments retain biological activity such
CC as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
CC AAW02747-W02999 for examples of polypeptide fragments). The polypeptide
CC fragments can be used in compositions for treating subjects suffering
CC from a pathology related to a GPR abnormality e.g. a psychotic disorder
CC such as schizophrenia. (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 412 AA;

Alignment Scores:

Pred. No.:	1.04e-18	Length:	412
Score:	361.50	Matches:	80
Percent Similarity:	51.73%	Conservative:	129
Best Local Similarity:	19.80%	Mismatches:	140
Query Match:	14.54%	Indels:	55
DB:	2	Gaps:	5

US-09-837-344-41 (1-1482) x AAW03626 (1-412)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||:::::   :::   |||:::|||         :::|||:::   :::|||:::
Db         33 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 52

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::   :::   |||::::::         :::|||:::   :::|||::::::   :::
Db         53 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 72

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
               |||:::|||         :::|||:::   :::|||::::::   :::   |||:::
Db         73 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 92

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||         :::|||:::   :::|||::::::   :::   |||:::|||
Db         93 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 112

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::|||:::   :::|||::::::   :::   |||::::::         :::|||
Db        113 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 132

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::   :::|||::::::   :::   |||:::|||         :::|||:::   :::
Db        133 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 152

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||::::::   :::   |||:::|||         :::|||:::   :::|||::::::
Db        153 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 172

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::   |||:::|||         :::|||:::   :::|||::::::   :::
Db        173 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 192

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::|||         :::|||:::   :::|||::::::   :::   |||:::|||
Db        193 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 212

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                    :::|||:::   :::|||::::::   :::   |||:::|||
Db        213 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 232

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::|||:::   :::|||::::::   :::   |||:::   :::   :::   ::::::
Db        233 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                    |||   ::::::|||
Db        253 GluGluGluGluGluGlu---------------------------------------- 259

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                          |||         |||   |||
Db        260 --------------GluGluGluGluGluGluGlu----------------------- 266

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
                        ::::::                  :::   :::|||
Db        267 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 282

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::         |||                     |||         :::     :::
Db        283 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 302

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    |||||:::::::::         :::   |||:::::|||::::::
Db        303 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 322

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::
Db        323 GluGlu---------------------------------------------------- 324

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                       |||:::|||         |||   ::::::         :::||||||
Db        325 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 340

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||::::::|||:::   :::|||   :::       |||   ::::::      |||:::
Db        341 GluGluGluGluGluGluGluGluGluGluGluTyrValPhePheGluGlu---GlnGlu 359

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::|||   ||||||      :::||||||::::::         :::   |||   |||
Db        360 AspGluIleIleGlyPheGlyGlnGluLeuLysAsn------ProGlnGluGluThrLeu 377

Qy       1201 GATGATTTAGAT 1212
                    |||
Db        378 GlnAlaPheAsp 381
```

```
RESULT 16
AAY96255
ID   AAY96255 standard; protein; 1162 AA.
XX
AC   AAY96255;
XX
DT   12-SEP-2003  (revised)
DT   11-SEP-2000  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;
KW   latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KW   Human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;
KW   Kaposi's sarcoma; primary effusion lymphoma; PEL;
KW   human immunodeficiency virus; HIV; multicentric Castleman's disease.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Domain          14. .17
FT                   /note= "nuclear localisation signal, NLS"
FT   Domain          64. .70
FT                   /note= "nuclear localisation signal, NLS"
FT   Region          320. .429
FT                   /note= "acidic repeat region"
FT   Region          430. .549
FT                   /note= "Gln, Glu, Pro-rich region"
FT   Region          550. .589
FT                   /note= "Gln, Glu, Pro, Arg-rich region"
FT   Region          590. .759
FT                   /note= "Gln, Glu, Asp-rich region"
FT   Region          760. .840
FT                   /note= "Gln, Glu-rich region"
XX
PN   WO200029626-A1.
XX
PD   25-MAY-2000.
XX
PF   19-NOV-1999;   99WO-US027508.
XX
PR   19-NOV-1998;   98US-00109422.
PR   21-APR-1999;   99US-00298568.
XX
PA   (KIEF/) KIEFF E D.
PA   (BALL/) BALLESTAS M E.
PA   (KAYE/) KAYE K M.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2000-387829/33.
DR   N-PSDB; AAA30290.
XX
PT   Treating or preventing a disease associated with rhodino virus infection
PT   in a mammal which includes Kaposi's Sarcoma and Primary Effusion
PT   Lymphoma.
XX
PS   Disclosure; Fig 7; 70pp; English.
XX
CC   The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC   (KSHV) latency-associated nuclear antigen (LANA). KSHV is also known as
```

```
CC   Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or gamma-2
CC   herpes virus class. The LANA protein is necessary for the efficient
CC   persistence of rhadino virus DNA in mammalian cells. Persistent rhadino
CC   virus infection is implicated in a variety of diseases e.g. Kaposi's
CC   Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
CC   Castleman's disease. In addition, KS is a common malignancy in HIV
CC   patients. KSHV persists in host cells in a latent form. One of the few
CC   genes expressed from the latent viral DNA is LANA. LANA associates with
CC   both human chromosomes and with the rhadino virus cis-acting element
CC   (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC   "tied" to the host chromosomes. This allows the viral DNA to persist in
CC   the host cell. The present sequence may be used to screen and identify
CC   molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC   with the latency cycle of this virus. Potential antiviral treatments for
CC   the above mentioned diseases may therefore be based on LANA deregulation.
CC   (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:              1.81e-18      Length:         1162
Score:                  359.00        Matches:        94
Percent Similarity:     49.44%        Conservative:   127
Best Local Similarity:  21.03%        Mismatches:     110
Query Match:            14.44%        Indels:         116
DB:                     3             Gaps:           12

US-09-837-344-41 (1-1482) x AAY96255 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||:::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          :::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                ||:::::      |||:::   |||::::::|||         ||||||:::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::             :::      ||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||:::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::          :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu---------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ---------------------------------------------------------- 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ---------------------------------------------------------- 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                      |||||||||::::::          :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                          ::::::   :::
Db        838 ---------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||          |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                 |||:::::::::::|||       |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              ::::::   ::::::   |||:::   :::          :::          ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 17
AAY58500
ID   AAY58500 standard; protein; 1162 AA.
XX
AC   AAY58500;
XX
DT   06-AUG-2003  (revised)
DT   10-APR-2000  (first entry)
XX
DE   HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW   HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen; antigen.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 96
FT                   /label= unknown
XX
PN   WO9961909-A2.
XX
PD   02-DEC-1999.
XX
PF   26-MAY-1999;   99WO-US011407.
XX
```

PR 26-MAY-1998; 98US-0086695P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pau C;
XX
DR WPI; 2000-097142/08.
XX
PT New methods and compositions for the detection of human herpesvirus.
XX
PS Claim 2; Page 59-62; 68pp; English.
XX
CC Sequences AAY58480-Y58532 represent immunogenic polypeptides derived from
CC human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC biological sample, and detecting the formation of a complex between the
CC peptide and the antibody. The presence of a peptide-antibody complex
CC indicates the presence of human herpesvirus 8. The detection of HHV8
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC HHV8-specific antibodies are useful therapeutically when for the passive
CC immunisation of a human against HHV8 infection, thereby reducing HHV8
CC related disease. The detection assays are highly specific, sensitive and
CC accurate. Early detection and treatment of Kaposi's sarcoma could
CC diminish the severity of symptoms related to AIDS and the sensitive
CC techniques could reduce erroneous characterisations of skin disorders.
CC Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC immunoblots and enzyme immunosassays lack the sensitivity and accuracy
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages of
CC the assays are that reproducible results are obtained and the method is
CC suitable for rapid throughput and screening of samples economically.
CC (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.81e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	49.44%	Conservative:	127
Best Local Similarity:	21.03%	Mismatches:	110
Query Match:	14.44%	Indels:	116
DB:	3	Gaps:	12

US-09-837-344-41 (1-1482) x AAY58500 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||       |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                 |||::::::      |||:::   |||::::::|||        ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::            :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu---------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ---------------------------------------------------------- 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ---------------------------------------------------------- 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                       |||||||||::::::            :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                              ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||          |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                 |||:::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              ::::::   ::::::   |||:::   :::          :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937
```

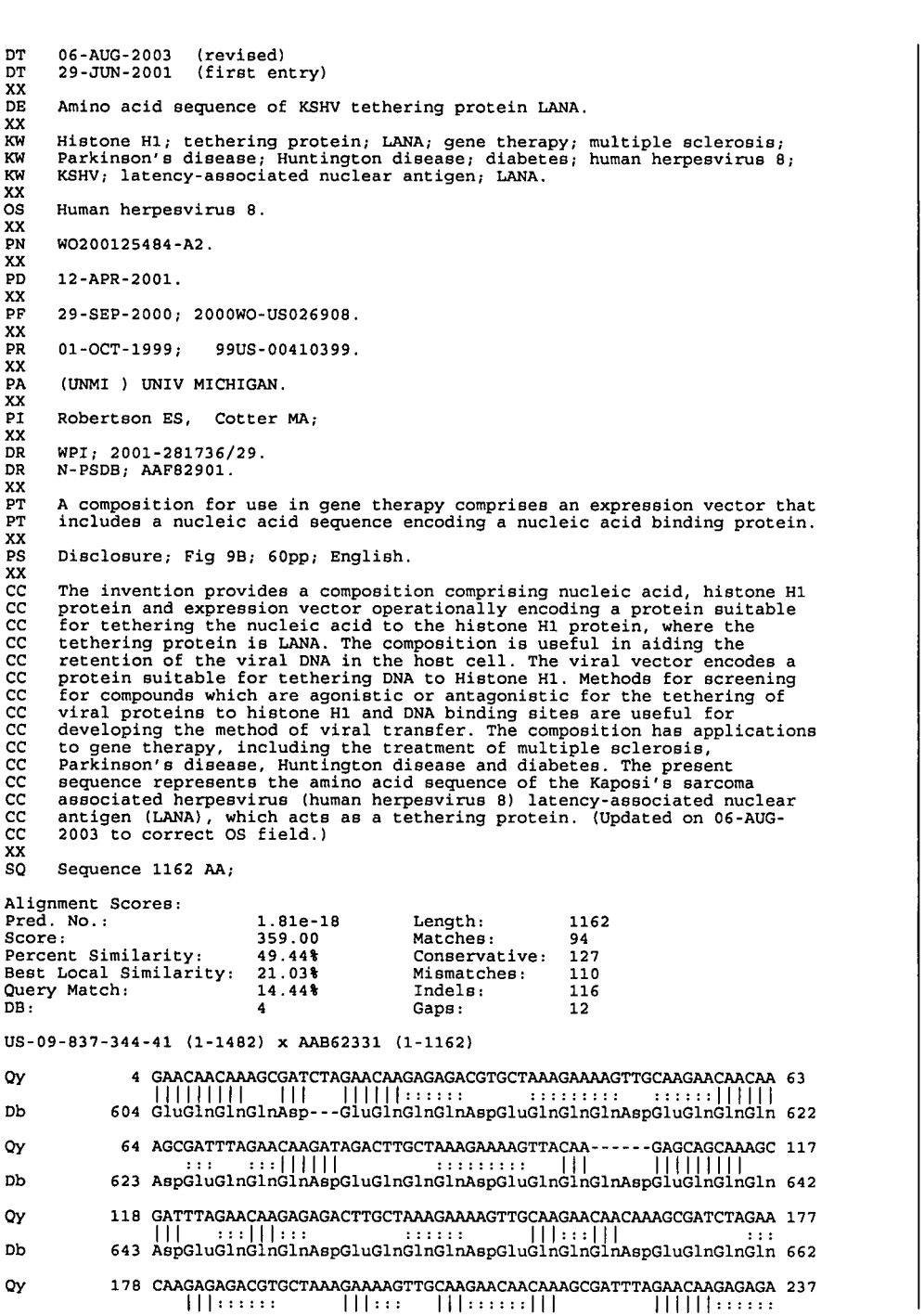
RESULT 18
AAB62331
ID AAB62331 standard; protein; 1162 AA.
XX
AC AAB62331;
XX

DT 06-AUG-2003 (revised)
DT 29-JUN-2001 (first entry)
XX
DE Amino acid sequence of KSHV tethering protein LANA.
XX
KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW KSHV; latency-associated nuclear antigen; LANA.
XX
OS Human herpesvirus 8.
XX
PN WO200125484-A2.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US026908.
XX
PR 01-OCT-1999; 99US-00410399.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Robertson ES, Cotter MA;
XX
DR WPI; 2001-281736/29.
DR N-PSDB; AAF82901.
XX
PT A composition for use in gene therapy comprises an expression vector that
PT includes a nucleic acid sequence encoding a nucleic acid binding protein.
XX
PS Disclosure; Fig 9B; 60pp; English.
XX
CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to Histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the amino acid sequence of the Kaposi's sarcoma
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-
CC 2003 to correct OS field.)
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.81e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	49.44%	Conservative:	127
Best Local Similarity:	21.03%	Mismatches:	110
Query Match:	14.44%	Indels:	116
DB:	4	Gaps:	12

US-09-837-344-41 (1-1482) x AAB62331 (1-1162)



```
Qy       1315 AGTTTGTATGATGAGCATATT 1335
                  :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 19
ABB05621
ID   ABB05621 standard; protein; 1162 AA.
XX
AC   ABB05621;
XX
DT   25-APR-2002  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA protein.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
KW   KSHV terminal repeat; rhadino virus cis acting element; episome;
KW   primary effusion lymphoma; latency-associated nuclear antigen;
KW   gene therapy; gene transfer.
XX
OS   Human herpesvirus 8.
XX
PN   US6322792-B1.
XX
PD   27-NOV-2001.
XX
PF   21-APR-1999;   99US-00298568.
XX
PR   19-NOV-1998;   98US-0109422P.
XX
PA   (KIEF/) KIEFF E D.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2002-153769/20.
DR   N-PSDB; ABA93487.
XX
PT   System for episomal retention of plasmids in mammalian cells, useful in
PT   gene therapy, comprises rhadinoviral LANA and RVCAE sequences.
XX
PS   Disclosure; Fig 7; 27pp; English.
XX
CC   The present invention describes a system (A) for maintaining a plasmid as
CC   an episome in mammalian cells, comprising the rhadinoviral sequence LANA
CC   (latency-associated nuclear antigen) of 3489 base pairs (see ABA93487,
CC   S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC   (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC   present in the plasmid. Also describes is a method for maintaining a
CC   closed circular DNA in a cell by expressing (S1) in the cells and having
CC   (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC   particularly used in gene therapy (or other gene transfer applications)
CC   that uses mammalian cells in which LANA is expressed. (A) improves
CC   persistence of gene therapy vectors in cells. The present sequence
CC   represents Kaposi's sarcoma-associated herpesvirus (KSHV, also called
CC   human herpesvirus 8) LANA protein, which is used in the exemplification
CC   of the present invention
XX
SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:              1.81e-18      Length:         1162
Score:                  359.00        Matches:        94
Percent Similarity:     49.44%        Conservative:   127
Best Local Similarity:  21.03%        Mismatches:     110
Query Match:            14.44%        Indels:         116
DB:                     5             Gaps:           12

US-09-837-344-41 (1-1482) x ABB05621 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::     :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||          :::::::::   |||       |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||           :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::       |||:::   |||::::::|||          ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::       ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   :::::::||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::          :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::          :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::          :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu---------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ---------------------------------------------------------- 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ---------------------------------------------------------- 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                      |||||||||::::::            :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                        ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||  |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||             |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
```

```
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
Db        931 GluThrValGluGluProIle 937

RESULT 20
ADJ65096
ID   ADJ65096 standard; protein; 1162 AA.
XX
AC   ADJ65096;
XX
DT   20-MAY-2004  (first entry)
XX
DE   HHV8 latency-associated nuclear antigen, LANA.
XX
KW   HHV8; latency-associated nuclear antigen; LANA;
KW   primary effusion lymphoma; PEL; virucide; gene therapy; KSHV;
KW   Kaposi's sarcoma-associated herpesvirus; episome; RVCAE;
KW   rhadinovirus cis-acting element; rhadinovirus infection.
XX
OS   Human herpesvirus 8.
XX
PN   US2004037847-A1.
XX
PD   26-FEB-2004.
XX
PF   28-JUN-2001; 2001US-00894273.
XX
PR   19-NOV-1998;   98US-0109422P.
PR   21-APR-1999;   99US-00298568.
XX
PA   (KIEF/) KIEFF E D.
PA   (BALL/) BALLESTAS M E.
PA   (KAYE/) KAYE K M.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2004-191011/18.
DR   N-PSDB; ADJ65095.
XX
PT   Assays for compounds that modulate rhadino virus LANA action in trans on
PT   a unit of rhadino virus DNA to mediate efficient episome persistence,
PT   comprises detecting whether the compound inhibits expression of a gene
PT   resident on that episome.
XX
PS   Disclosure; SEQ ID NO 2; 28pp; English.
XX
CC   The invention relates to an assay for determining whether a compound
CC   modulates LANA (latency-associated nuclear antigen) from KSHV (Kaposi's
CC   sarcoma-associated herpesvirus, also known as HHV8, human herpesvirus 8)
CC   mediated persistence in a mammalian cell of a DNA episome having RVCAE
CC   (rhadinovirus cis-acting element) by detecting whether the compound
CC   inhibits expression of a gene resident on that episome. Also included are
CC   maintaining a closed circular DNA in a mammalian cell in which LANA is
CC   expressed (by using RVCAE as the cis-acting origin and maintenance
CC   sequence) and maintaining a plasmid as an episome in a mammalian cell
CC   (comprising: (a) Expressing LANA in the cell; and (b) having RVCAE
CC   resident on the plasmid). Compounds identified as interfering with the
CC   interaction between LANA and RVCAE are useful in treating or preventing
CC   rhadino virus infection (e.g. primary effusion lymphoma (PEL)). LANA acts
CC   in trans on a 0.8 Kb KSHV TR (terminal repeat) unit to mediate efficient
CC   episome persistence. The present sequence represents KSHV LANA.
XX

SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:              1.81e-18      Length:          1162
Score:                  359.00        Matches:           94
Percent Similarity:     49.44%        Conservative:     127
Best Local Similarity:  21.03%        Mismatches:       110
Query Match:            14.44%        Indels:           116
DB:                     8             Gaps:              12

US-09-837-344-41 (1-1482) x ADJ65096 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
Db        809 GlnGluLeuGluGluGlnGluGlnGlu------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db        817 ------------------------------------------------------------ 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
Db        817 ------------------------------------------------------------ 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
```

```
                    |||||||||:::::             :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                      ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::         |||             |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                |||:::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                ::::::   ::::::   |||:::   :::            :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGluGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 21
ADV68154
ID   ADV68154 standard; protein; 1162 AA.
XX
AC   ADV68154;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA protein.
XX
KW   screening; cytostatic; virucide; gene therapy;
KW   latency-associated nuclear antigen; tumor; viral infection;
KW   herpesvirus type 8 infection.
XX
OS   Human herpesvirus 8.
XX
PN   US2004248081-A1.
XX
PD   09-DEC-2004.
XX
PF   11-JUL-2002; 2002US-00194046.
XX
PR   01-OCT-1999;   99US-00410399.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Robertson ES;
XX
DR   WPI; 2005-038492/04.
DR   N-PSDB; ADV68153.
XX
PT   New composition having a latency-associated nuclear antigen (LANA)
PT   nucleic acid, useful for detecting compounds for treating tumors and
PT   viral infections.
XX
PS   Disclosure; SEQ ID NO 2; 51pp; English.
XX
CC   The invention relates to a composition comprising a latency-associated
CC   nuclear antigen (LANA) nucleic acid having any of 14 fully defined
CC   sequences of 10-61 bp given in the specification (ADV68159-ADV68172). The
CC   methods and compositions of the present invention are useful for
CC   detecting compounds that are agonistic or antagonistic for the binding of
CC   viral genetic material to genomic host DNA, in particular for detecting
CC   compounds for treating tumor and viral infections. This sequence
CC   corresponds to the LANA protein from the Kaposi's sarcoma-associated
CC   herpesvirus (also called human herpesvirus 8). The LANA protein tethers
CC   the viral DNA to the human chromosomal structural protein histone H1.
XX
SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:                1.81e-18        Length:         1162
Score:                    359.00          Matches:        94
Percent Similarity:       49.44%          Conservative:   127
Best Local Similarity:    21.03%          Mismatches:     110
Query Match:              14.44%          Indels:         116
DB:                       9               Gaps:           12

US-09-837-344-41 (1-1482) x ADV68154 (1-1162)

Qy           4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
               |||||||||   |||   |||||::::::      :::::::::  ::::::||||||
Db         604 GluGlnGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy          64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                  :::   :::||||||          :::::::::   |||      |||||||||
Db         623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy         118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
               |||   :::|||:::          ::::::      |||:::|||            :::
Db         643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy         178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                  |||::::::      |||:::   |||::::::|||          ||||||::::::
Db         663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy         238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                    :::::::::   ::::::||||||   :::   :::||||||          :::
Db         683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy         298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
               :::      ||||||||||||   :::   ||||||:::         ::::::
Db         703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy         358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
               |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db         723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy         418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
               |||   |||   :::|||:::            :::      ||||||||||   :::
Db         743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy         478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                  ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db         759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy         538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
               :::         :::::::::|||:::||||||:::   :::||||||::::::
Db         779 Gln---------GluGlnGluLeuGluGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy         598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                  :::::::::|||:::||||||:::::::::||||||::::::            :::
Db         794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy         658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
               :::   ||||||   :::::::::
Db         809 GlnGluLeuGluGluGlnGluGlnGlu-------------------------------- 817

Qy         718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db         817 ------------------------------------------------------------ 817

Qy         778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db         817 ------------------------------------------------------------ 817
```

```
Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu------------------------------------ 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                  |||||||||::::::          :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                          ::::::   :::
Db        838 ---------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::         |||          |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                |||:::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              ::::::   ::::::   |||:::   :::         :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937
```

RESULT 22
AAB44672
ID AAB44672 standard; protein; 207 AA.
XX
AC AAB44672;
XX
DT 12-FEB-2001 (first entry)
XX
DE Gene 33 human secreted protein homologous amino acid sequence #137.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW food additive; preservative; chromosome identification; cancer;
KW immune disorder; cardiovascular disorder; neurological disease;
KW wound healing; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200058339-A2.
XX
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US007440.
XX
PR 26-MAR-1999; 99US-0126503P.
PR 17-DEC-1999; 99US-0172409P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-594637/56.
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
PS Disclosure; Page 399-400; 410pp; English.
XX
CC The polynucleotide sequences given in AAC79681 to AAC79730 encode the
CC human secreted proteins given in AAB44596 to AAB44645. AAB44646 to
CC AAB44693 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC can also be used as a food additive or preservative to increase or
CC decrease storage capabilities. The polynucleotides are useful for
CC chromosome identification. They are also useful as probes for diagnosing
CC a disorder related to the female reproductive system, particularly breast
CC and/or ovary cancer. They are also useful in the gene therapy of breast
CC and ovarian cancer. The nucleic acids, protein, antibodies, agonists and
CC antagonists from the present invention are useful in the diagnosis,
CC treatment and prevention of: cancer; immune disorders; cardiovascular
CC disorders; wound healing; neurological diseases; and infectious diseases.
CC AAC79672 to AAC79680 and AAB44595 represent sequences used in the
CC exemplification of the present invention
XX
SQ Sequence 207 AA;

Alignment Scores:

Pred. No.:	4.24e-17	Length:	207
Score:	339.50	Matches:	75
Percent Similarity:	72.25%	Conservative:	89
Best Local Similarity:	33.04%	Mismatches:	34
Query Match:	13.66%	Indels:	29
DB:	3	Gaps:	12

US-09-837-344-41 (1-1482) x AAB44672 (1-207)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::|||:::      |||||||||
Db          1 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGluGlnGlnAspGluGlnGln 19

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                 |||   :::|||   :::      :::|||:::   :::||||||:::
Db         20 GlnAspGluGlnGlnGlnGlnAspGluGlnGluGlnGlnGluGluGlnGluGlnGlnGlu 39

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||||||||:::      :::|||:::   ||||||||||||   :::   |||||||||
Db         40 GluGlnGluGln---GlnGluGluGlnGluGlnGluGlnGlnGluGluGlnGluGlnGlu 58

Qy        184 AGACGTGCTAAAGAA---AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
                    :::|||   :::|||:::||||||:::   :::||||||::::::
Db         59 LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 76

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::::::::|||:::||||||:::   :::||||||:::            ::::::
Db         77 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln---------GluGln 92

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::|||:::||||||:::   :::||||||::::::         :::::::::|||:::
Db         93 GluLeuGluGluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGlu 109

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||:::   :::||||||::::::         :::::::::|||:::||||||:::
Db        110 GluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGlu 126

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::||||||::::::         :::::::::|||:::||||||:::   :::|||
Db        127 GlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeu 143
```

```
Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        144 GluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln 160

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        161 ---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln--------- 174

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
Db        175 GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------GluGlnGluGln 192

Qy        661 AATTTAGAAAGAAAAAAGGAA 681
Db        193 GluLeuGluGluValGluGlu 199

RESULT 23
AAR70491
ID   AAR70491 standard; protein; 562 AA.
XX
AC   AAR70491;
XX
DT   27-AUG-2003  (revised)
DT   19-DEC-1995  (first entry)
XX
DE   Leucocytozoan protozoa structural protein epitope.
XX
KW   leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
KW   leucocytozoanosis; treatment.
XX
OS   Leucocytozoon.
XX
PN   JP07089995-A.
XX
PD   04-APR-1995.
XX
PF   10-SEP-1993;   93JP-00226078.
XX
PR   10-SEP-1993;   93JP-00226078.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (NISS-) NISSEIKEN KK.
XX
DR   WPI; 1995-167252/22.
DR   N-PSDB; AAQ87587.
XX
PT   Immune inducing polypeptide against Leucocytozoan protozoa - useful in
PT   production of vaccines for treatment of leucocytozoanosis in fowl.
XX
PS   Claim 1; Page 12-14; 20pp; Japanese.
XX
CC   AAR70491-93 are polypeptides having a whole or partial epitope of a
CC   structural protein of Leucocytozoan protozoa encoded by AAQ87587-89. The
CC   polypeptides and DNA encoding them are useful in the production of
CC   vaccines for the treatment of leucocytozoanosis of fowl. (Updated on 27-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 562 AA;

Alignment Scores:
Pred. No.:              1.04e-16        Length:         562
Score:                  335.00          Matches:        118
Percent Similarity:     51.07%          Conservative:   120
Best Local Similarity:  25.32%          Mismatches:     146
Query Match:            13.48%          Indels:         82
DB:                     2               Gaps:           22

US-09-837-344-41 (1-1482) x AAR70491 (1-562)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db         44 GluGlnGluGluGluGlnGluGlnGluGluGlnGluGlnGluIleValGluGluGln 63

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db         64 GluGlnAspGluGluGln---------GluGluGluAspGluGluGluGluGluGlu 80

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db         81 LysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        101 GlnAspGluGluGluGluGluGluGluGlnAspGluAspGluGluGluGluGluAsp 120

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---CTTGCTAAA 297
Db        121 GluGluGluGluGluGluGluGluGlnAsnGluGluGluGlnAsnGluAspGluGlnAsn 140

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db        141 GluAspGluGlnAsnGluGluGluGluGluGluGluGluGluGlnGlnGluGlnAsp 160

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db        161 GluGluGluGlnAspGluGluGluGlnAspGluGluGluGluGluGluGlnGluGlu 180

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCT----------------------------AAA 450
Db        181 GlnAspGluGluGlnGluGluValTyrAlaGluLysGluAsnGluAspGluGluLysLys 200

Qy        451 GAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGTGCTAAA 501
Db        201 GluLysGluGluGluGlnGluAspGluLysIleTyrValGluLysGluLysAspGluGlu 220

Qy        502 ---------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
Db        221 ValTyrValGluLysGluGlnGluHisPheLysGluLeuGluLysGlu---ValLeuGlu 239

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
Db        240 GluLysValGluSerGluHisLysAspAspLysLysGluThrLysIleLysLysHisGlu 259

Qy        613 CAAGAGCAGCAAAGAGATTTA-----------GAACAAAGGAAGGCTGATACGAAAAAA 660
Db        260 ValGluGluGluArgGluIleGluIleLysGlyGluSerGluGluIleAspGluLysLys 279

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db        280 Ser---GluValLysGluGlnAsnGlyGlnValGlnGlyLys----------------- 292

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA------CCA 774
Db        293 --------------------------ThrAsnGluGlnLysGluTyrIleGlnAspPro 303

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT 834
Db        304 IleTrpSerSerIleLeuArgTyrMetIleAsnLysArgThrGluThr------------ 319

Qy        835 ATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT------GAAGGACGAAGG 888
Db        320 ------GluLysThrAsnGlnGlnAsnTyrAsnIleLysValThrAsnGluSerLysSer 337

Qy        889 GATATACATAAAGGA---------CATCTTGAAGAAAAGAAAGATGGTTCAATA---AAA 936
Db        338 GluIleAsnSerThrThrAsnIlePheGluAspGluAsnLysAsnSerSerLeuLysLys 357

Qy        937 CCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
Db        358 ProGluLeuAsnGluAlaThrArgGlyArgAsnSerGlyProValValGluGluVal--- 376

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA--- 1053
Db        377 ---------ValGluGluThrProVal------ValGluGluGluMetValGluGluThr 391

Qy       1054 ------TATGACGATTCATTAATAGATGAA------------GAAGAAGAT--------- 1086
```

```
                    ::::::   :::|||:::|||           ||||||:::
Db        392 ProValValGluGluGluMetIleGluGluThrProValValGluGluGluMetIleGlu 411

Qy       1087 --------------GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
                         :::||||||:::   |||      |||||||||:::   |||
Db        412 GluThrProValValGluGluAspValValGluGluThrProIleValGluGluAsp--- 430

Qy       1132 TTCCAAGATGAAGAAAACATA------GGAATTTATAAAGAACTA------------GAA 1173
                ::::|||   |||:::         :::   :::|||                |||
Db        431 ValValGluGluThrAsnValValGluGluValValGluGluThrProValValGluGlu 450

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
              ::::::||||||:::         ::::::::::::   |||
Db        451 GluMetIleGluGluThrProValValGluGluValValGluGluThrProValValGlu 470

Qy       1234 GAAGAATTATCTGAAGAA 1251
              ||||||:::   ||||||
Db        471 GluGluValValGluGlu 476

RESULT 24
AAB18165
ID   AAB18165 standard; protein; 1192 AA.
XX
AC   AAB18165;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:22.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US026796.
XX
PR   05-NOV-1998;   98US-0107131P.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection.
XX
PS   Disclosure; Page 60-63; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC   (II) are useful for the development of vaccines against P. falciparum
CC   infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC   immunogens comprising the sequences of (I), are useful in the detection
CC   of infection with P. falciparum. Furthermore, (I) (especially when they
CC   are rifins or secreted or membrane proteins) can aid the identification
CC   of drugs to treat or prevent P. falciparum infection, or they can be used
CC   to identify drug resistance in P. falciparum. Sequencing of the
CC   Plasmodium chromosome 2 and the subsequent identification of proteins
CC   encoded by it will help to expand our understanding of parasite biology,
CC   a process hampered by the complexity of the parasitic lifecycle, and
CC   provide new targets for vaccine and drug development. Parasite resistance
CC   to drugs and mosquito resistance to insecticides have led to a resurgence
CC   of malaria in many parts of the world, and there is a pressing need for
CC   vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC   represent nucleotide and protein sequences given in the present
CC   invention, but which are not specifically mentioned within the
CC   specification
XX
SQ   Sequence 1192 AA;

Alignment Scores:
Pred. No.:                1.75e-16      Length:          1192
Score:                    332.50        Matches:         112
Percent Similarity:       46.85%        Conservative:    126
Best Local Similarity:    22.05%        Mismatches:      193
Query Match:              13.37%        Indels:          77
DB:                       3             Gaps:            16

US-09-837-344-41 (1-1482) x AAB18165 (1-1192)

Qy         10 CAAAGCGATCTAGAACAAGAGAGA--------------------------------CGT 36
              ||||||      :::|||||||||                                 :::
Db        296 GlnSerLysTyrLysGlnGluArgIleGluIleLeuAspAspAsnGlyLysGluLeuLys 315

Qy         37 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA---------GATAGACTT 87
              :::   :::   ::::::|||::::::     :::|||:::              :::
Db        316 SerHisLysAsnIleLysGluGluLysGlyGlyIleGluLysThrAspThrThrAsnIle 335

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 147
              |||      |||:::::::::::::::   :::   :::   |||:::      ::::::
Db        336 AlaAspIleLysIleLysLysGluGluArgGluThrLysAspGluLys------GluLys 353

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207
                 :::|||:::          |||::::::      :::      :::|||
Db        354 AsnIleGlnGlnLeuValLysAspValGlnLeuIleLysValGlyGluGluThrLysAsp 373

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
              :::::::::   |||   |||                :::::      |||:::
Db        374 AspGluLysGluAspLysGluGlyThrAspAspGluGluAspThrAspAspGluGluAsp 393

Qy        268 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 327
              :::|||   |||:::|||         :::|||         ::::::::::::
Db        394 ThrAspAspGluGluAspThrAspAspGluGluAspThrSerAspGluGluThrThrGly 413

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::||||||   :::   :::   ::::::   |||::::::::::::
Db        414 AspGlnGluAsnLysGluGluThrGluValAspGluLysLysThrGlu------------ 429

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
              :::|||:::|||:::|||:::|||   :::   :::   |||::::::::::
Db        430 ---LysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGluGlu 448

Qy        448 AAAGAAAAG---TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 504
              ||||||      :::|||   :::   |||   |||:::      :::   ||||||
Db        449 SerGluLysAspLysGluGluSerGluLysAspLysGluGluSerGluLysAspLysGlu 468

Qy        505 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 564
              |||   :::|||   :::          |||   |||:::      :::          :::
Db        469 LysThrGluGluAspGluGluLysThrGluAspGluLysGlyThrGluValTyrLysLys 488

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAA 624
              |||          |||   :::::::|||:::                :::      :::  :::
Db        489 GluThrAspValAspGluLysLysGluLysGlyGluTyrGlyGluGlyThrAspAspGlu 508

Qy        625 AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACAT 684
              |||   |||:::   :::   |||   :::      :::|||   ||||||
Db        509 GluAspLysGluLysGluGluAspAspGluGluThrLysValGluGluLysLysThrGlu 528

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCA 744
              |||
Db        529 LysAsp---------------------------------------------------- 530

Qy        745 TCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA 804
                       |||  |||                :::      :::|||
Db        531 ---------GluGluGly-----------------ThrAspTyrGluGluAspThrAsp 541
```

```
Qy        805 GGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGA---GAATCT 861
                 :::   :::|||   :::               :::||||||   |||   |||
Db        542 AspSerAspLysAspGluGluThrLysValGluGluLysLysThrGluArgAspGluGlu 561

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA 921
                |||         |||         ::::::   |||         |||:::   :::
Db        562 GluThrGluGluAspGluLysGluThrGluValGluLysLysLysThrGluLysAspGlu 581

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                :::|||:::         |||       :::|||          |||::::::
Db        582 GluGlyThrAspTyrGluGluAspThrAspAspSerAspLysAspValGluThrGluVal 601

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAG 1035
                ||||||           |||       |||   :::   ::::::         :::
Db        602 GluGluThr-----------AspAlaGluAspLysGluGluAsnGluGluGlyThrAsp 617

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT---GATGAA 1092
                ||||||                :::       |||   |||:::|||||||||   :::|||
Db        618 AspGluGluAspLysValGluGluThrAspLeuAspAspGlnGluGluAspGlyGluGlu 637

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
                |||   ::::::   |||         :::   |||:::   :::|||::::::
Db        638 AspLysGluAspAspLysGluLysAspLysGluAspAspLysGluAspAspLysGlu--- 656

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA--------------------AAT 1191
                      |||:::   ||||||      ||||||                     :::
Db        657 ---------LysAspLysGluAspAspLysGluLysTyrLysGluAspAspLysGluAsp 673

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
                :::       ||||||   ::::::       |||   :::   |||:::       |||:::
Db        674 AspLysGluAspAspLysGluLysAspLysGluAspAsnLysGluLysAspLysGluAsp 693

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
                    :::|||   |||:::   :::|||   ||||||            |||   :::|||
Db        694 AsnLysGluLysAspLysGluAspAspLysGluLysAsp---------LysGluAspAsp 710

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
                |||         |||       |||       :::   ||||||:::   |||||||||
Db        711 LysGluLysAspLysGluAspAsnLys------GluLysAspLysGluAspAsnLysGlu 728

Qy       1372 AAGGAAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                |||:::|||                 |||             |||       :::   :::
Db        729 LysAspLysGluAspAspLysGluLysHis-----------AspLysHisValArgArg 744

Qy       1432 ATCGTGGATGAGTTATCTGAAGAT 1455
                |||       ::::::   :::|||
Db        745 IleLysLysLysMetLysAspAsp 752

RESULT 25
ABB58657
ID   ABB58657 standard; protein; 2274 AA.
XX
AC   ABB58657;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 2763.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02760.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 2763; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2274 AA;

Alignment Scores:
Pred. No.:                2.46e-16      Length:          2274
Score:                    331.00        Matches:         134
Percent Similarity:       44.30%        Conservative:    134
Best Local Similarity:    22.15%        Mismatches:      167
Query Match:              13.31%        Indels:          170
DB:                       4             Gaps:            27

US-09-837-344-41 (1-1482) x ABB58657 (1-2274)

Qy           1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTG 51
                 :::|||::::::      |||       ||||||       |||          ||||||:::
Db        1385 LysGluArgGluGluSerLeuArgMetGluArgGluGluLysLeuLysGluGluLysIle 1404

Qy          52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
                 :::|||::::::          :::::::::|||   :::||||||:::::|||:::
Db        1405 LysGluLysGlu---------ArgGluGluLysLeuLysGluGluLysIleLysGluLys 1421

Qy         112 CAAAGCGAT------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
                 |||   :::      |||::::::|||:::|||   :::::::::   :::|||:::
Db        1422 GlnArgGluGluLysLeuLysGluGluLysLeuLysGluLysGluArgGluGluArgMet 1441

Qy         166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA--------- 216
                :::   |||:::|||   :::||||||||||||   :::|||::::::
Db        1442 LysGluLysGluArgGluGluLysAlaLysGluLysGlnArgGluGluLysLeuArgGlu 1461

Qy         217 ---------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                        :::   |||:::|||   :::   ||||||||||||:::|||::::::
Db        1462 GluLysIleLysGluLysGluArgGluGluLysLeuLysGluLysLeuArgGluGluLys 1481

Qy         268 ---AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
                   :::   |||::::::         :::::::::   :::|||:::      :::
Db        1482 IleLysGluLysGluLysGluGluLysLeuArgLysGluArgGluGluLysMetArgGlu 1501

Qy         325 TTAGAACAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAACAA 369
                   |||:::|||   :::   |||              ||||||::::::|||::::::
Db        1502 LysGluArgGluGluLysIleLysGluLysGluArgValGluLysIleLysGluLysGlu 1521

Qy         370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                   :::   :::   ::::::      :::||||||||||:::|||::::::      |||
Db        1522 ArgGluGluLysLeuLysLysGluLysGluGluLysLeuLysGluLysGluGluLeuLeu 1541

Qy         430 GAA-------------------------CAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
                :::                           :::|||:::|||   :::::::::   |||
Db        1542 LysLysLysGluLysGluGluLysGluArgGluGluLysLeuLysGluLysGluArgGln 1561

Qy         463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
```

```
Db       1562 GluLysLeuLysGluLysGluArgGluGluLysLeuLysArgGluThrGluGluArgGln 1581

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGAT 579

Db       1582 ArgGluLysGluArgGluGluLysLeuLysGluLysGluArgAlaGluLysLeuLysAsp 1601

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAGCAGCAAAGAGAT 630

Db       1602 LeuGluLysGluValLysLeuLysGluLysGluGluGlnLeuLysGluLysGluLysGlu 1621

Qy        631 TTA---------GAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681

Db       1622 LeuLysLeuLysGluLysLysGluLysAspLysValLysGluLysGluLysSerLeuGlu 1641

Qy        682 CATGGAGATATATTA----------------------------------GCAGAGGAT 705

Db       1642 SerGluLysLeuLeuIleSerAlaThrValSerAsnProTrpArgArgValValGluAsp 1661

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATA---GAACTTCCATCAGAAAATGAACGTGGA 762

Db       1662 ThrProProLysLeu------ProAlaValGlnAspTyrProSerLeuGlyLysLys--- 1678

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA------------------ 804

Db       1679 ---------ProThrLysAlaSer---ProGluLysLysArgAspGluLysLeuLeuPro 1694

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATATCT---------------------------- 834

Db       1695 GlyLeuThrThrProProLysGluValAsnAspThrPheGluAspPheLeuSerGlyLeu 1714

Qy        835 ------------------------------------------------ATAATAGAAAAA 846

Db       1715 LysProLeuGluAlaLeuProProLeuProAlaLeuGluProLeuGluValLysGluAsp 1734

Qy        847 ACAAATAGAGAATCTATTACA---ACAAATGTTGAAGGA---------------------- 882

Db       1735 SerLysLysGluSerValSerLeuIleAsnPheGluSerProLeuGlnGluThrAlaSer 1754

Qy        883 ---CGAAGGGATATACAT---------AAAGGACATCTTGAAGAAAAG------------ 918

Db       1755 IleProArgLysIleSerProProProArgGlyPheThrGluGlnAsnLeuIleLeuAla 1774

Qy        919 AAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA------------TCTGCTGAC 966

Db       1775 LeuCysGlySerLeuHisTyrGluAsnGluGlnGluArgIleArgGluThrGluAlaGlu 1794

Qy        967 ATACAAAATCAT-----------------------------ACATTAGAGACAGTA--- 993

Db       1795 ValGlnProGluValValThrThrProGluProValTyrIleThrLeuAspThrLeuAla 1814

Qy        994 --------------------------AATATTTCTGATGTTAATGATTTTCAAATA--- 1023

Db       1815 LeuGlnLysSerThrSerThrSerAsnAsnSerSerGlySerGluGluIleValValMet 1834

Qy       1024 ----------------------------AGTAAGTATGAGGATGAAATAAGTGCTGAA 1053

Db       1835 GluGluProValLysLysLeuSerLysLysHisLysArgLeuLysGlnLeuGlnLeuGln 1854

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113

Db       1855 GlnGlnGlnGlnGlnArgGluArgGluProAspAspGluGluLeuArgProLeuIleSer 1874

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173

Db       1875 IleSerMetCysAspSerGlnLeuAspValProAspIleProVal-------------- 1889

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233

Db       1890 -----------------ArgAlaMetProAspAspAspGluGlyValGluGluGluPro 1903

Qy       1234 GAAGAATTATCTGAAGAAAAAATA---------------------------------- 1257
```

```
Db       1904 GluThrGluProGluGlnSerLeuProGluAspLeuLeuHisPheGlySerSerGlyVal 1923

Qy       1258 ----------------------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293

Db       1924 AlaThrAsnAlaThrThrThrAspSerGluGlyProValProAlaThrThrSerAspAsp 1943

Qy       1294 AATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGAT 1353

Db       1944 AsnValValTyrGlySerSerThrSerAsnAlaAlaProHisLysLeuLysThrLys 1963

Qy       1354 AAGCAGGTTAATAAG 1368

Db       1964 LysLeuGluHisLys 1968

RESULT 26
ADP29884
ID   ADP29884 standard; protein; 2990 AA.
XX
AC   ADP29884;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human secreted protein SEQ ID #651.
XX
KW   Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW   cancer; inflammatory; immune; human secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO2004035732-A2.
XX
PD   29-APR-2004.
XX
PF   28-AUG-2003; 2003WO-US026780.
XX
PR   29-AUG-2002; 2002US-0406576P.
PR   29-AUG-2002; 2002US-0406579P.
PR   29-AUG-2002; 2002US-0406585P.
PR   29-AUG-2002; 2002US-0406588P.
PR   29-AUG-2002; 2002US-0406608P.
PR   29-AUG-2002; 2002US-0406611P.
PR   29-AUG-2002; 2002US-0406612P.
PR   29-AUG-2002; 2002US-0406616P.
PR   29-AUG-2002; 2002US-0406640P.
PR   29-AUG-2002; 2002US-0406642P.
PR   29-AUG-2002; 2002US-0406646P.
PR   29-AUG-2002; 2002US-0406653P.
PR   29-AUG-2002; 2002US-0406655P.
PR   29-AUG-2002; 2002US-0406666P.
PR   17-SEP-2002; 2002US-0410946P.
PR   17-SEP-2002; 2002US-0410947P.
PR   17-SEP-2002; 2002US-0410948P.
PR   17-SEP-2002; 2002US-0410949P.
PR   17-SEP-2002; 2002US-0410953P.
PR   17-SEP-2002; 2002US-0410957P.
PR   17-SEP-2002; 2002US-0410958P.
PR   17-SEP-2002; 2002US-0410959P.
PR   17-SEP-2002; 2002US-0410960P.
PR   17-SEP-2002; 2002US-0410961P.
PR   17-SEP-2002; 2002US-0410962P.
PR   17-SEP-2002; 2002US-0411019P.
PR   17-SEP-2002; 2002US-0411022P.
PR   17-SEP-2002; 2002US-0411023P.
PR   17-SEP-2002; 2002US-0411024P.
PR   17-SEP-2002; 2002US-0411032P.
PR   17-SEP-2002; 2002US-0411035P.
PR   17-SEP-2002; 2002US-0411037P.
PR   17-SEP-2002; 2002US-0411041P.
PR   17-SEP-2002; 2002US-0411045P.
PR   17-SEP-2002; 2002US-0411046P.
PR   17-SEP-2002; 2002US-0411048P.
PR   17-SEP-2002; 2002US-0411052P.
PR   17-SEP-2002; 2002US-0411055P.
```

```
PR   17-SEP-2002; 2002US-0411073P.
PR   17-SEP-2002; 2002US-0411082P.
PR   17-SEP-2002; 2002US-0411101P.
PR   17-SEP-2002; 2002US-0411111P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT,  Chu K,  Lee E,  Hestir K,  Beaurang PA,  Behrens D;
PI   Halenbeck RF,  Huang MM,  Kothakota S,  Haishan L,  Linnemann T;
PI   Pierce K,  Wang Y,  Wong JGP,  Wu G,  Zhang H;
XX
DR   WPI; 2004-348438/32.
XX
PT   New nucleic acid molecule for diagnosing, preventing or treating diseases
PT   such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT   genetic, bacterial and viral diseases.
XX
PS   Claim 1; SEQ ID NO 1882; 428pp; English.
XX
CC   The present invention relates to an isolated nucleic acid molecule
CC   encoding a polypeptide which is believed to be cytostatic,
CC   antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC   composition and methods are useful for diagnosing, preventing and
CC   treating diseases such as proliferative (e.g. cancer), inflammatory,
CC   immune, metabolic, genetic, bacterial and viral diseases. The present
CC   sequence represents a human secreted protein. The present sequence is
CC   available on WIPOWEB and is not in the specification.
XX
SQ   Sequence 2990 AA;

Alignment Scores:
Pred. No.:                2.77e-16        Length:          2990
Score:                    330.50          Matches:         156
Percent Similarity:       46.43%          Conservative:    91
Best Local Similarity:    29.32%          Mismatches:      128
Query Match:              13.29%          Indels:          157
DB:                       8               Gaps:            34

US-09-837-344-41 (1-1482) x ADP29884 (1-2990)

Qy          1 CAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAA---------AAG 48
              :::||||||:::      |||   |||||||||   |||   :::|||         :::
Db       1490 ArgGluGlnGluGluArgLeuArgGluGlnGluGluArgGlnArgGluGlnGluAspArg 1509

Qy         49 TTGCAAGAACAACAAAGCGATTTA------GAACAAGATAGACTTGCT------AAAGAA 96
              |||   ||||||:::      |||         :::::::::||||||         :::|||
Db       1510 LeuHisGluGlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluGlu 1529

Qy         97 AAGTTACAAGAGCAGCAAAGCGATTTA------GAACAAGAGAGACTTGCT------AAA 144
              :::|||   ||||||:::      |||         :::|||||||||          :::
Db       1530 ArgLeuCysGluGlnGluGluArgLeuArgGluHisGluGluArgLeuCysGluGlnGlu 1549

Qy        145 GAAAAGTTGCAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAA------ 195
              |||:::|||:::||||||:::      |||   |||||||||   |||   :::
Db       1550 GluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluGluArgLeuArgGluHis 1569

Qy        196 ---GAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGT--------- 240
                 |||:::|||   ||||||:::      |||   |||||||||   |||
Db       1570 GluGluArgLeuCysGluGlnGluGluArgLeuCysGluGlnGluGluArgLeuHisGlu 1589

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGATAGACTTGCTAAA 297
                 :::|||:::|||:::||||||:::      |||   ||||||:::
Db       1590 GlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGlu------------ 1605

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA------ 348
              |||:::|||:::|||   :::      |||   |||||||||   |||   :::
Db       1606 GluArgLeuArgGluHisGluGluArgLeuCysGluGlnGluGluArgLeuArgGluHis 1625

Qy        349 ---GAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGT--------- 393
                 |||:::|||   ||||||:::      |||   |||||||||   |||
Db       1626 GluGluArgLeuCysGluGlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGlu 1645

Qy        394 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACTTGCTAAA 450
                 :::|||:::|||:::||||||:::      |||   |||||||||:::|||
Db       1646 GlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluLysLeuProGly 1665

Qy        451 GAAAAG------TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 504
              ::::::      |||:::|||   :::      ||||||||||||||||||   :::|||
Db       1666 GlnGluArgLeuLeuGluGluValGluLysLeuLeuGluGlnGluArgArgGlnGluGlu 1685

Qy        505 AAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAACAA 537
              :::                         |||:::|||   :::      |||||||||
Db       1686 GlnGluArgLeuLeuGluArgGluArgLeuLeuGluGluValGluLysLeuLeuGluGln 1705

Qy        538 GAGAGACGTGCTAAAGAAAAG--------------------------TTGCAAGAACAA 570
              ||||||:::   :::|||:::                         |||:::|||
Db       1706 GluArgGlnGlnGluGluGlnGluArgLeuLeuGluArgGluArgLeuLeuGluGluVal 1725

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAGCAGCAA 624
              :::      ||||||||||||||||||         :::|||:::|||   |||::::::
Db       1726 GluLysLeuLeuGluGlnGluArgArgGlnGluGluGlnGluArgLeuLeuGluArgGlu 1745

Qy        625 AGA---------GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              |||         ::::::|||:::         :::|||   :::   |||||||||
Db       1746 ArgLeuLeuAspGluValGluGluLeuLeuAspGluThrLeuGlnGluLeuGluArgLeu 1765

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
              :::|||                                                  :::
Db       1766 ArgGlu----------------------------------------------------Leu 1768

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
              |||         |||   ||||||         :::      ::::::         |||
Db       1769 GluArgLeuArgGluLeuGluArgMetLeuGluLeuGlyTrpGluAlaLeuTyrGluGln 1788

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGA 855
                  |||         |||         |||      |||:::   |||
Db       1789 ------ArgAlaGluProArg--------------SerGlyPheGluGluLeuAsnAsn 1801

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
              |||:::   :::|||                  :::   :::         |||
Db       1802 GluAsnLysSerThr--------------LeuGlnLeuGluGlnGlnValLysGluLeu 1816

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
              ||||||   |||   :::|||                  |||   |||      :::
Db       1817 LysLysSerGlyGluLeuLysGluThrValThrSerAspProSerLysLysMet------ 1834

Qy        976 CATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAG 1035
                                                  ::::::         ::::::
Db       1835 ------------------------------------TrpGluProIleValPheLys 1841
```

```
Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT--------- 1086
Db       1842 GluLysLeuThrMetLys---ThrAspSerLeuMetGluGluLysLeuGluCysSerLeu 1860

Qy       1086 ------------------------------------------------------------ 1086

Db       1861 TrpCysCysLeuSerAsnProSerIleProGlyArgCysCysValLeuGluArgArgIle 1880

Qy       1087 --------------GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
                             |||                 ||||||      |||:::
Db       1881 ValProTrpMetGlnGlnGluSerTyrSerSerSerSerProIleTrpSerValAspSer 1900

Qy       1132 TTCCAAGATGAAGAAAACATA---GGAATTTATAAAGAACTAGAAGATTTGATAGAG--- 1185
                 ||||||   |||:::      :::   :::   |||||||||      |||
Db       1901 ------AspGluProAsnLeuThrSerValLeuGluArgLeuGluAspThrLysGluAsn 1918

Qy       1186 ---------------------------------------AAAAATGAAAATTTAGATGAT 1206
                                                     |||   |||   ||||||
Db       1919 SerSerProLeuPheAsnCysProProProThrThrSerLysThrGluIleLeuAspSer 1938

Qy       1207 TTAGAT--------------------GAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 :::                    |||   ||||||      :::||||||
Db       1939 CysAsnCysSerProSerGluSerValArgGlyAlaGluLysGlyAspGlnGluLeuLeu 1958

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
                  ||||||:::   :::|||:::|||||||
Db       1959 ---AlaAsnIleLysGlnAlaGluLysHisGluLys 1969
```

RESULT 27
AAW03627
ID AAW03627 standard; peptide; 360 AA.
XX
AC AAW03627;
XX
DT 25-MAR-2003 (revised)
DT 17-DEC-1996 (first entry)
XX
DE Human follicle stimulating hormone GPR N-terminal sequence.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
XX
OS Homo sapiens.
XX
PN US5508384-A.
XX
PD 16-APR-1996.
XX
PF 09-SEP-1993; 93US-00118270.
XX
PR 10-SEP-1992; 92US-00943236.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Schuster DI, Murphy RB;
XX
DR WPI; 1996-208785/21.
XX
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
PT treating schizophrenia.
XX
PS Disclosure; Fig 8B(2); 184pp; English.
XX
CC Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomegaloviral and other GPR proteins. The peptides AAW03578-
CC W03651 represent the N-terminal fragments of the above proteins. The
CC receptor proteins were used to design polypeptides, pref. based on the
CC transmembrane domains, for use in G-protein coupled receptor ligand
CC binding assays. The polypeptide fragments retain biological activity such
CC as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
CC AAW02747-W02999 for examples of polypeptide fragments). The polypeptide
CC fragments can be used in compositions for treating subjects suffering
CC from a pathology related to a GPR abnormality e.g. a psychotic disorder
CC such as schizophrenia. (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 360 AA;

Alignment Scores:

Pred. No.:	3.6e-16	Length:	360
Score:	327.50	Matches:	70
Percent Similarity:	49.60%	Conservative:	116
Best Local Similarity:	18.67%	Mismatches:	132
Query Match:	13.17%	Indels:	57
DB:	2	Gaps:	3

US-09-837-344-41 (1-1482) x AAW03627 (1-360)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||:::::   :::   |||:::|||          :::|||:::   :::|||:::
Db         33 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 52

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::   :::   |||::::::         :::|||:::   :::|||::::::   :::
Db         53 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 72

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                 |||:::|||          :::|||:::   :::|||::::::   :::   |||:::
Db         73 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 92

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||          :::|||:::   :::|||::::::   :::   |||:::|||
Db         93 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 112

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::|||:::   :::|||::::::   :::   |||::::::          :::|||
Db        113 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 132

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::   :::|||::::::   :::   |||:::|||          :::|||:::   :::
Db        133 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 152

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||::::::   :::   |||:::|||          :::|||:::   :::|||::::::
Db        153 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 172

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::   |||:::|||          :::|||:::   :::|||::::::   :::
Db        173 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 192

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::|||          :::|||:::   :::|||::::::   :::   |||:::|||
Db        193 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 212

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                       ::|||:::   :::|||::::::   :::   |||:::|||
Db        213 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 232

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::|||:::   :::|||::::::   :::   |||:::   :::   :::   ::::::
Db        233 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                    |||   ::::::|||
Db        253 GluGluGluGluGluGlu---------------------------------------- 259

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Db        259 ---------------------------------------------------------- 259

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Db        259 ---------------------------------------------------------- 259
```

```
Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::      |||          |||      :::      :::
Db        260 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 279

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                  ||||||:::::::::         :::   |||::::::|||::::::
Db        280 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 299

Qy        961 GCTGACATACAAAATCATACATTA---GAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                 |||          :::      ::::::::::   |||:::
Db        300 ValAspTyrMetThrGlnAlaArgGlyGlnArgSerSerLeuAlaGluAspAsnGluSer 319

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                |||:::      |||          |||:::|||      |||   :::|||
Db        320 SerTyrSerArgGlyPheAspMetThrTyrThrGluPheAspTyrAspLeuCysAsnGlu 339

Qy       1078 GAAGAAGAT---------GATGAAGACTTAGACGAATTTAAGCCT 1113
                 |||                      |||   |||   |||
Db        340 ValValAspValThrCysSerProLysProAspAlaPheAsnPro 354
```

RESULT 28
AAY29039
ID AAY29039 standard; protein; 611 AA.
XX
AC AAY29039;
XX
DT 24-SEP-1999 (first entry)
XX
DE T. gondii immunogenic protein.
XX
KW Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
KW T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
KW Toxoplasma oocyst.
XX
OS Toxoplasma gondii.
XX
PN WO9932633-A1.
XX
PD 01-JUL-1999.
XX
PF 18-DEC-1998; 98WO-US027137.
XX
PR 19-DEC-1997; 97US-00994825.
XX
PA (HESK-) HESKA CORP.
XX
PI Milhausen MJ, Lutz SB, Ng RK;
XX
DR WPI; 1999-418930/35.
DR N-PSDB; AAX91242.
XX
PT New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT infection caused by this microorganism.
XX
PS Claim 29; Page 227-229; 381pp; English.
XX
CC The invention provides isolated Toxoplasma gondii nucleic acids that
CC encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC immunogenic proteins and antibodies to the proteins can be used to
CC inhibit T. gondii oocyst shedding in a cat due to infection with T.
CC gondii. They can be used for preventing T. gondii infection and for
CC preventing the spread of T. gondii infection. They can also be used for
CC detecting T. gondii infection. The detection method can be used to detect
CC parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC such as Cryptosporidium oocysts and Toxoplasma oocysts
XX
SQ Sequence 611 AA;

Alignment Scores:
Pred. No.: 7.64e-16 Length: 611
Score: 323.50 Matches: 78
Percent Similarity: 64.66% Conservative: 83

Best Local Similarity: 31.33% Mismatches: 69
Query Match: 13.01% Indels: 19
DB: 2 Gaps: 9

US-09-837-344-41 (1-1482) x AAY29039 (1-611)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||      |||:::
Db        323 ArgGluArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::|||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||      :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                 |||:::|||||||||            |||   :::   :::|||:::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||       |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::          |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGlu 462

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
              ||||||      |||:::   :::|||          |||:::   :::   |||:::
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495
              |||:::   :::||||||   :::|||:::|||   :::   |||:::||||||||
Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546
                 :::||||||   :::|||:::|||   :::   ::::::||||||||
Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606
                 :::||||||   :::::::::|||   :::       :::||||||:::      |||
Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::|||::::::|||:::   :::||||||:::   :::   |||      |||
Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690
                 |||   ::::::||||||   :::
Db        563 ValGluLeuGlnArgGluHisAlaGlu 571
```

RESULT 29
AAU25510
ID AAU25510 standard; protein; 611 AA.
XX
AC AAU25510;
XX
DT 17-DEC-2001 (first entry)
XX
DE T. gondii immunogenic protein POC-23.
XX
KW Immunogenic protein; oocyst; faeces; enteric apicomplexa oocyst;
KW Cryptosporidium oocyst; Toxoplasma oocyst; Giardia cyst; vaccine;
KW oocyte shedding.
XX
OS Toxoplasma gondii.

XX
PN US2001014447-A1.
XX
PD 16-AUG-2001.
XX
PF 18-DEC-1998; 98US-00216393.
XX
PR 19-DEC-1997; 97US-00994825.
XX
PA (MILH/) MILHAUSEN M J.
XX
PI Milhausen MJ;
XX
DR WPI; 2001-529100/58.
DR N-PSDB; AAS42565.
XX
PT Detecting parasite oocysts or cysts in feces, comprises eluting DNA from
PT sample into aqueous solution by heating, amplifying DNA with primers
PT specific for oocysts or cysts being detected, and detecting amplification
PT product.
XX
PS Example 2; Page 91-93; 188pp; English.
XX
CC The invention relates to detection of parasite oocysts or cysts in a
CC faeces sample comprising contacting the sample with a solid support,
CC drying and then washing the sample with an aqueous wash solution, adding
CC an aqueous elution solution and eluting DNA from the sample by heating
CC and amplifying by PCR oocyst/cyst-specific DNA and detecting the
CC amplification products. The method is useful for detecting parasite
CC oocysts e.g., enteric apicomplexa oocysts such as Cryptosporidium oocysts
CC or Toxoplasma oocysts, or for detecting parasite cysts e.g. Giardia
CC cysts. The method is also useful for developing vaccines to prevent
CC oocyte shedding in cats. The present sequence represents an immunogenic
CC protein from Toxoplasma gondii
XX
SQ Sequence 611 AA;

Alignment Scores:

Pred. No.:	7.64e-16	Length:	611
Score:	323.50	Matches:	78
Percent Similarity:	64.66%	Conservative:	83
Best Local Similarity:	31.33%	Mismatches:	69
Query Match:	13.01%	Indels:	19
DB:	4	Gaps:	9

US-09-837-344-41 (1-1482) x AAU25510 (1-611)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||      |||:::
Db        323 ArgGluArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||      :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                 |||:::|||||||||               |||   :::   :::|||::::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||      |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::         |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGlu 462

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
              ||||||      |||:::   :::|||         |||:::   :::   |||:::
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495
              |||:::   :::||||||   :::|||:::|||   :::   |||:::|||||||||
Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546
                 :::||||||   :::|||:::|||   :::   :::::::|||||||||
Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606
                 :::||||||   ::::::::::|||   :::       :::||||||:::      |||
Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::|||::::::|||:::   :::|||||||:::   :::   |||      |||
Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690
                 |||   ::::::||||||   :::
Db        563 ValGluLeuGlnArgGluHisAlaGlu 571
```

RESULT 30
ADG17131
ID ADG17131 standard; protein; 611 AA.
XX
AC ADG17131;
XX
DT 26-FEB-2004 (first entry)
XX
DE T. gondii protein #39.
XX
KW Toxoplasma gondii; oocyst shedding; genetic vaccine; vaccine.
XX
OS Toxoplasma gondii.
XX
PN US2003194393-A1.
XX
PD 16-OCT-2003.
XX
PF 17-DEC-2002; 2002US-00321856.
XX
PR 19-DEC-1997; 97US-00994825.
PR 18-DEC-1998; 98US-00216393.
XX
PA (MILH/) MILHAUSEN M J.
XX
PI Milhausen MJ;
XX
DR WPI; 2003-899768/82.
DR N-PSDB; ADG17130.
XX
PT Novel immunogenic Toxoplasma gondii proteins useful for inhibiting oocyst
PT shedding by cats infected with Toxoplasma gondii.
XX
PS Claim 6; SEQ ID NO 81; 198pp; English.
XX
CC The invention relates to an isolated Toxoplasma gondii protein. The
CC protein is useful for inhibit oocyst shedding by cats infected with
CC Toxoplasma gondii. The protein is useful for preventing or ameliorating
CC diseases caused by infection with T. gondii. The nucleic acid can be used
CC as genetic vaccine which encodes the protein. The protein and the nucleic
CC acid are used as diagnostic reagents for detection of T. gondii
CC infection. The present sequence is used in the exemplification of the
CC invention.
XX
SQ Sequence 611 AA;

Alignment Scores:

Pred. No.:	7.64e-16	Length:	611
Score:	323.50	Matches:	78
Percent Similarity:	64.66%	Conservative:	83
Best Local Similarity:	31.33%	Mismatches:	69
Query Match:	13.01%	Indels:	19
DB:	7	Gaps:	9

US-09-837-344-41 (1-1482) x ADG17131 (1-611)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||       |||:::
Db        323 ArgGluArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||       :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                |||:::|||||||||             |||   :::   :::|||::::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||      |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::          |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluArgGlu 462

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
              ||||||      |||:::   :::|||          |||:::   :::   |||:::
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495
              |||:::   :::||||||   :::|||:::|||   :::   |||:::|||||||||
Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546
                 :::||||||   :::|||:::|||   :::   ::::::|||||||||
Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606
              :::||||||   :::::::::|||   :::      :::||||||:::      |||
Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::|||::::::|||:::   :::||||||:::   :::   |||      |||
Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690
                 |||   ::::::||||||   :::
Db        563 ValGluLeuGlnArgGluHisAlaGlu 571
```

Search completed: December 21, 2005, 10:45:00
Job time : 322.224 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 51.3343 Seconds
(without alignments)
4773.622 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093212_22815/app_query.fasta_1.3740
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_161_@runat_21122005_093212_22815 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1549	62.3	316	2	US-08-098-327E-31	Sequence 31, Appl
2	1549	62.3	316	2	US-08-462-625-31	Sequence 31, Appl
3	603	24.3	117	2	US-08-098-327E-20	Sequence 20, Appl
4	603	24.3	117	2	US-08-462-625-20	Sequence 20, Appl
5	554	22.3	107	2	US-08-098-327E-19	Sequence 19, Appl
6	554	22.3	107	2	US-08-462-625-19	Sequence 19, Appl
7	376	15.1	1898	1	US-08-056-200-94	Sequence 94, Appl
8	376	15.1	1898	1	US-08-800-644-94	Sequence 94, Appl
9	376	15.1	1898	2	US-09-538-092-1280	Sequence 1280, Ap
10	359	14.4	1162	1	US-08-728-323A-2	Sequence 2, Appli
11	359	14.4	1162	2	US-09-298-568-2	Sequence 2, Appli
12	359	14.4	1162	2	US-09-410-399-2	Sequence 2, Appli
13	359	14.4	1162	2	US-09-894-273-2	Sequence 2, Appli
14	336	13.5	608	2	US-09-270-767-32937	Sequence 32937, A
15	336	13.5	608	2	US-09-270-767-48154	Sequence 48154, A
16	323.5	13.0	611	2	US-09-216-393B-81	Sequence 81, Appl
17	319	12.8	1972	2	US-08-875-435B-3	Sequence 3, Appli
18	317.5	12.8	1937	2	US-09-538-092-918	Sequence 918, App
19	316	12.7	1180	2	US-09-543-681A-6436	Sequence 6436, Ap
20	315.5	12.7	1564	2	US-10-144-198-2	Sequence 2, Appli
21	315.5	12.7	1564	2	US-10-144-198-4	Sequence 4, Appli
22	315	12.7	1960	2	US-09-538-092-1077	Sequence 1077, Ap
23	315	12.7	1960	2	US-09-949-016-10872	Sequence 10872, A
24	314	12.6	1976	2	US-09-538-092-1078	Sequence 1078, Ap
25	309	12.4	1972	2	US-08-875-435B-4	Sequence 4, Appli
26	306.5	12.3	1942	2	US-09-949-016-8135	Sequence 8135, Ap
27	305	12.3	1939	2	US-09-538-092-915	Sequence 915, App
28	305	12.3	1939	2	US-09-949-016-11104	Sequence 11104, A
29	304.5	12.2	1939	2	US-09-949-016-6925	Sequence 6925, Ap
30	303.5	12.2	924	2	US-09-248-796A-18798	Sequence 18798, A
31	303	12.2	1939	2	US-09-310-187A-1	Sequence 1, Appli
32	303	12.2	1939	2	US-09-538-092-917	Sequence 917, App
33	302	12.1	1959	2	US-09-949-016-8134	Sequence 8134, Ap
34	301	12.1	1857	2	US-09-917-254-91	Sequence 91, Appl
35	301	12.1	1972	2	US-09-538-092-1084	Sequence 1084, Ap
36	301	12.1	1984	2	US-09-949-016-7111	Sequence 7111, Ap
37	301	12.1	1984	2	US-09-949-016-7112	Sequence 7112, Ap
38	301	12.1	1984	2	US-09-949-016-7113	Sequence 7113, Ap
39	299.5	12.0	1886	2	US-08-938-105-3	Sequence 3, Appli
40	299.5	12.0	1935	2	US-09-538-092-916	Sequence 916, App
41	299	12.0	3878	2	US-09-914-259-11	Sequence 11, Appl
42	298.5	12.0	1944	2	US-09-949-016-10929	Sequence 10929, A
43	298	12.0	735	2	US-10-164-595-80	Sequence 80, Appl
44	298	12.0	784	2	US-10-164-595-79	Sequence 79, Appl
45	298	12.0	843	2	US-10-164-595-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-08-098-327E-31
; Sequence 31, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; APPLICANT: DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/098,327E
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620

```
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-31

Alignment Scores:
Pred. No.:              1.09e-117      Length:         316
Score:                  1549.00        Matches:        314
Percent Similarity:     99.37%         Conservative:   0
Best Local Similarity:  99.37%         Mismatches:     2
Query Match:            62.31%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x US-08-098-327E-31 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 2
US-08-462-625-31
; Sequence 31, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-31

Alignment Scores:
Pred. No.:              1.09e-117      Length:         316
Score:                  1549.00        Matches:        314
Percent Similarity:     99.37%         Conservative:   0
Best Local Similarity:  99.37%         Mismatches:     2
Query Match:            62.31%         Indels:         0
DB:                     2              Gaps:           0
```

```
US-09-837-344-41 (1-1482) x US-08-462-625-31 (1-316)

Qy          13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy          73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy         133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy         193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy         253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy         313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy         373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy         433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
               |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db         141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy         493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy         553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy         613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy         673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy         733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy         793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy         853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy         913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 3
US-08-098-327E-20
; Sequence 20, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-20

Alignment Scores:
Pred. No.:              4.03e-41        Length:         117
Score:                  603.00          Matches:        117
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            24.26%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-41 (1-1482) x US-08-098-327E-20 (1-117)

Qy         610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy         670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy         730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy         790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy         850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100
```

```
Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 4
US-08-462-625-20
; Sequence 20, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-20

Alignment Scores:
Pred. No.:              4.03e-41      Length:          117
Score:                  603.00        Matches:         117
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            24.26%        Indels:          0
DB:                     2             Gaps:            0

US-09-837-344-41 (1-1482) x US-08-462-625-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 5
US-08-098-327E-19
; Sequence 19, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-19

Alignment Scores:
Pred. No.:              3.71e-37      Length:          107
Score:                  554.00        Matches:         107
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            22.28%        Indels:          0
DB:                     2             Gaps:            0
```

```
US-09-837-344-41 (1-1482) x US-08-098-327E-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 6
US-08-462-625-19
; Sequence 19, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-19

Alignment Scores:
Pred. No.:              3.71e-37        Length:         107
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            22.28%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-41 (1-1482) x US-08-462-625-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 7
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/056,200
;      FILING DATE:  30-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  94:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-056-200-94

Alignment Scores:
Pred. No.:              1.44e-22      Length:          1898
Score:                  376.00        Matches:         137
Percent Similarity:     47.39%        Conservative:    117
Best Local Similarity:  25.56%        Mismatches:      174
Query Match:            15.12%        Indels:          108
DB:                     1             Gaps:            24

US-09-837-344-41 (1-1482) x US-08-056-200-94 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||:::::|||||   :::             |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                 ||||||   |||||||||   :::         :::|||:::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC------------------ 219
              ||||||||||   :::         :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                  |||:::|||||||||   :::         :::||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::                |||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                 ::::::||||||   :::|||:::   :::|||:::              :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 ---------------------------------GATTTAGAACAAGAG--------- 540
                                                  ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||      |||:::      |||:::                    :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---------------------- 615
                 |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
                  ||||||::::::   |||||||||:::      |||   :::         |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT-------------- 717
              |||   |||   |||              :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                        ::::::   |||   |||                ::::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::             |||       |||                |||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATATCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::       :::|||:::       :::                        :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||       :::         |||::::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::            :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                        :::                 ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                    |||             |||:::   |||||||||::::::         :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                        :::      :::   :::|||            |||:::
Db        935 ------------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::       |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::              :::      :::|||             |||      ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 8
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
```

```
;       STATE:  CA
;       COUNTRY:  U.S.A.
;       ZIP:  92660
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/800,644
;       FILING DATE:  14-FEB-1997
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/056,200
;       FILING DATE:  30-APR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fedrick, Michael F.
;       REGISTRATION NUMBER:  36,799
;       REFERENCE/DOCKET NUMBER:  NIH054.001A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (714) 760-0404
;       TELEFAX:  (714) 760-9502
;   INFORMATION FOR SEQ ID NO:  94:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1898 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-800-644-94

Alignment Scores:
Pred. No.:                1.44e-22        Length:          1898
Score:                    376.00          Matches:         137
Percent Similarity:       47.39%          Conservative:    117
Best Local Similarity:    25.56%          Mismatches:      174
Query Match:              15.12%          Indels:          108
DB:                       1               Gaps:            24

US-09-837-344-41 (1-1482) x US-08-800-644-94 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||        |||||||| |||  :::::::::|||:::  |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::|||||   ::            |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              ||||||   |||||||||  :::         :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC------------------ 219
              ||||||||||||  :::         :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
              |||:::|||||||||  :::         :::||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::                |||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg------------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||  |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                  :::::||||||   :::|||:::   :::|||:::                :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   ::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluAlaArg 713

Qy        526 -----------------------------------GATTTAGAACAAGAG--------- 540
                                                  ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||       |||:::       |||:::                  :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
              ||||||::::::   |||||||||:::      |||   :::         |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT-------------- 717
              |||   |||   |||             :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||  |||                :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::         |||      |||                ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATATCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                    :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||      :::          |||::::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::          :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                  :::                 ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                |||          |||:::   |||||||||::::::        :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                  :::      :::   :::|||            |||:::
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::      |||  :::  :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::              :::      :::|||            |||      ||||||
Db        987 GluGluGlnLeuGluGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
```

```
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 9
US-09-538-092-1280
; Sequence 1280, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1280
;   LENGTH: 1898
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number Q07283
US-09-538-092-1280

Alignment Scores:
Pred. No.:                1.44e-22      Length:          1898
Score:                    376.00        Matches:          137
Percent Similarity:       47.39%        Conservative:     117
Best Local Similarity:    25.56%        Mismatches:       174
Query Match:              15.12%        Indels:           108
DB:                       2             Gaps:              24

US-09-837-344-41 (1-1482) x US-09-538-092-1280 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::        |||::::::||||||   :::            |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                 ||||||   |||||||||   :::         :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
              ||||||||||||   :::         :::|||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                 |||:::|||||||||   :::         :::||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::                |||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                 ::::::||||||   :::|||:::   :::|||:::               :::

Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
                 :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 -----------------------------------GATTTAGAACAAGAG--------- 540
                                                  ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||      |||:::     |||:::                     :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                 |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
                 ||||||::::::   |||||||||:::      |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
              |||   |||   |||             :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                          ::::::   |||   |||                :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::             |||       |||                  ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATATCTATAATAGAAAAAACAAATAGA---------------------GAATCTATTACA 867
              :::       :::|||:::      :::                        :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||       :::          |||:::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::              :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                        :::                  ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                 |||                |||:::   |||||||||::::::         :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                        :::      :::   :::|||             |||:::
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::       |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::            :::|||                |||       ||||||
Db        987 GluGluGlnLeuGluGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022
```

```
RESULT 10
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma-Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1162 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-728-323A-2

Alignment Scores:
Pred. No.:                3.2e-21        Length:          1162
Score:                    359.00         Matches:         94
Percent Similarity:       49.44%         Conservative:    127
Best Local Similarity:    21.03%         Mismatches:      110
Query Match:              14.44%         Indels:          116
DB:                       1              Gaps:            12

US-09-837-344-41 (1-1482) x US-08-728-323A-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::     :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::        ::::::      |||:::|||           :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||:::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   :::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::              :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::          :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::           :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ------------------------------------------------------------ 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ------------------------------------------------------------ 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::     :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                       |||||||||:::::              :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                          ::::::   :::
Db        838 ---------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||            |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                 |||:::::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              ::::::   ::::::      |||:::   :::                  :::   ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
```

```
Db        931 GluThrValGluGluProIle 937

RESULT 11
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT: Ballestas, Mary E.
;  APPLICANT: Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/298,568
;  CURRENT FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,422
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

Alignment Scores:
Pred. No.:                3.2e-21        Length:          1162
Score:                    359.00         Matches:         94
Percent Similarity:       49.44%         Conservative:    127
Best Local Similarity:    21.03%         Mismatches:      110
Query Match:              14.44%         Indels:          116
DB:                       2              Gaps:            12

US-09-837-344-41 (1-1482) x US-09-298-568-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||::::::|||          ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGlnGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
```

```
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::            :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu-------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ------------------------------------------------------------ 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ------------------------------------------------------------ 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                      |||||||||::::::            :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                        ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGluGlnGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||             |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
              |||:::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                 ::::::   ::::::   |||:::   :::            :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 12
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
;  APPLICANT:  Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION:  Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE:  UM-03778
;  CURRENT APPLICATION NUMBER: US/09/410,399
;  CURRENT FILING DATE:  1999-10-01
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2
```

Alignment Scores:

Pred. No.:	3.2e-21	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	49.44%	Conservative:	127
Best Local Similarity:	21.03%	Mismatches:	110
Query Match:	14.44%	Indels:	116
DB:	2	Gaps:	12

US-09-837-344-41 (1-1482) x US-09-410-399-2 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||           :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                 |||::::::      |||:::   |||::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||           :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::             :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::       :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::           :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu---------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ------------------------------------------------------------ 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ------------------------------------------------------------ 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                       |||||||||:::::::            :::   ||||||
```

```
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                                 :::::::   :::
Db        838 ------------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::          |||             |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                 |||:::::::::::::|||       |||:::       ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGluGluGluGluGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                 ::::::   ::::::   |||:::   :::             :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937
```

RESULT 13
US-09-894-273-2
; Sequence 2, Application US/09894273
; Patent No. 6756203
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/894,273
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/109,422
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

Alignment Scores:

Pred. No.:	3.2e-21	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	49.44%	Conservative:	127
Best Local Similarity:	21.03%	Mismatches:	110
Query Match:	14.44%	Indels:	116
DB:	2	Gaps:	12

US-09-837-344-41 (1-1482) x US-09-894-273-2 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||           :::
```

```
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
              |||::::::      |||:::  |||:::::::|||        ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
              :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::             :::      ||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
              ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::::::::|||:::||||||:::::::::||||||::::::         :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu-------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        817 ---------------------------------------------------------- 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ---------------------------------------------------------- 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                    |||||||||::::::              :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                      ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||   |||:::      :::                 ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::         |||         |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                  |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGlnGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
              |||::::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                    ::::::   ::::::   |||:::   :::         :::   ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                        :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 14
US-09-270-767-32937
; Sequence 32937, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32937 .
;   LENGTH: 608
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-32937

Alignment Scores:
Pred. No.:              2.14e-19        Length:         608
Score:                  336.00          Matches:        70
Percent Similarity:     64.94%          Conservative:   80
Best Local Similarity:  30.30%          Mismatches:     73
Query Match:            13.52%          Indels:         8
DB:                     2               Gaps:           4

US-09-837-344-41 (1-1482) x US-09-270-767-32937 (1-608)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::         |||:::|||||||||   :::|||:::   :::|||:::
Db        374 LysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArgLysGluGlu 393

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::|||      :::|||:::   :::|||::::::
Db        394 GluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGluGluArgArg 413

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                  |||:::   |||      :::::::::   :::|||::::::         |||:::
Db        414 LysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArgLysGluGlu 433

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ::::::   :::|||:::   :::|||   :::         |||:::::::::::
Db        434 GluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGluGluArgLys 452

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||:::|||:::   ::::::::::::         |||:::   |||      :::|||
Db        453 AlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArgArgGluGlu 472

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              |||   :::   ::::::         |||:::   |||:::   :::|||:::   :::
Db        473 LysArgArgArgGluGlu---------GluGluArgArgLysGluGluGluArgArgGlu 489

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTGCAAGAACAA 417
              |||::::::         |||:::|||||||||      :::|||:::   :::|||:::
Db        490 GluGluGluLysArgLysGluGluGluArgArgLysAspGluGluArgArgArgGluGlu 509

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              :::         |||:::||||||      ||||||:::   :::|||:::
Db        510 GluLysArgLysGluGluGluArgArgGluLysGluArgArgArgGluGluGlyLysArg 529

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                  |||:::|||||||||   ||||||:::   :::|||::::::         |||:::
Db        530 LysGluGluGluArgArgGluLysGluArgArgArgGluGluGluLysArgLysGluGlu 549
```

```
Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              |||||||||  ||||||:::     |||:::::      |||:::  ||||||
Db        550 GluArgArgGluLysGluArgArgAspGluGluArgArgArgGluGluGluArgArgArg 569

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA---------AGGAAGGCT 648
                 :::|||:::   :::|||::::::|||      |||:::         |||:::
Db        570 GluGluGluArgArgArgGluGluGluArgArgArgGluGluGluArgArgArgGluGlu 589

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::   ::::::      |||||||||:::|||
Db        590 GluArgArgArgGluGluGluArgLysArgGlu 600

RESULT 15
US-09-270-767-48154
; Sequence 48154, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48154
;   LENGTH: 608
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-48154

Alignment Scores:
Pred. No.:              2.14e-19        Length:         608
Score:                  336.00          Matches:        70
Percent Similarity:     64.94%          Conservative:   80
Best Local Similarity:  30.30%          Mismatches:     73
Query Match:            13.52%          Indels:         8
DB:                     2               Gaps:           4

US-09-837-344-41 (1-1482) x US-09-270-767-48154 (1-608)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::          |||:::|||||||||   :::|||:::   :::|||:::
Db        374 LysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArgLysGluGlu 393

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::|||      :::|||:::   :::|||::::::
Db        394 GluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGluGluArgArg 413

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                 |||:::   |||      :::::::::   :::|||::::::          |||:::
Db        414 LysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArgLysGluGlu 433

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||:::   :::|||:::   :::|||   :::         |||:::||||||:::
Db        434 GluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGluGluArgLys 452

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||:::|||:::   ::::::::::::          |||:::   |||       :::|||
Db        453 AlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArgArgGluGlu 472

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              |||   :::   ::::::          |||:::   |||:::   :::|||:::   :::
Db        473 LysArgArgArgGluGlu---------GluGluArgArgLysGluGluGluArgArgGlu 489

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTGCAAGAACAA 417
              |||::::::          |||:::|||||||||      :::|||:::   :::|||:::
Db        490 GluGluGluLysArgLysGluGluGluArgArgLysAspGluGluArgArgArgGluGlu 509

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              :::         |||:::||||||      ||||||:::   :::|||:::
Db        510 GluLysArgLysGluGluGluArgArgGluLysGluArgArgArgGluGluGlyLysArg 529

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 |||:::|||||||||   ||||||:::   :::|||::::::          |||:::
Db        530 LysGluGluGluArgArgGluLysGluArgArgArgGluGluGluLysArgLysGluGlu 549

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              |||||||||  ||||||:::     |||::::::          |||:::  ||||||
Db        550 GluArgArgGluLysGluArgArgAspGluGluArgArgArgGluGluGluArgArgArg 569

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA---------AGGAAGGCT 648
                 :::|||:::   :::|||::::::|||      |||:::         |||:::
Db        570 GluGluGluArgArgArgGluGluGluArgArgArgGluGluGluArgArgArgGluGlu 589

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::   ::::::      |||||||||:::|||
Db        590 GluArgArgArgGluGluGluArgLysArgGlu 600

RESULT 16
US-09-216-393B-81
; Sequence 81, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THE
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393B
;  CURRENT FILING DATE:  1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393B-81

Alignment Scores:
Pred. No.:              2.2e-18         Length:         611
Score:                  323.50          Matches:        78
Percent Similarity:     64.66%          Conservative:   83
Best Local Similarity:  31.33%          Mismatches:     69
Query Match:            13.01%          Indels:         19
DB:                     2               Gaps:           9

US-09-837-344-41 (1-1482) x US-09-216-393B-81 (1-611)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||      |||:::
Db        323 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||          :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                 |||:::|||||||||              |||   :::   :::|||::::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||       |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::          |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGlu 462
```

```
Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495
Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546
Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606
Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690
Db        563 ValGluLeuGlnArgGluHisAlaGlu 571

RESULT 17
US-08-875-435B-3
; Sequence 3, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
;  APPLICANT: Hasegawa, Kazuhide
;  APPLICANT:  Arakawa, Emi
;  APPLICANT:  Oda, Shoji
;  APPLICANT:  Matsuda, Yuzuru
;  APPLICANT:  Takahashi, Katsuhito
;  APPLICANT:  Sugahara, Michihiro
;  APPLICANT:  Ishiyama, Haruo
;  TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
;  TITLE OF INVENTION:  FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
;  TITLE OF INVENTION:  VECTOR DNA, MICROORGANSIM CARRYING THE RECOMBINANT DNA, AND
;  TITLE OF INVENTION:  AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
;  TITLE OF INVENTION:  RECOMBINANT DNA
;  FILE REFERENCE: 07898-013001
;  CURRENT APPLICATION NUMBER: US/08/875,435B
;  CURRENT FILING DATE:  1997-07-25
;  PRIOR APPLICATION NUMBER: PCT/JP96/00134
;  PRIOR FILING DATE: 1996-01-25
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-08-875-435B-3

Alignment Scores:
Pred. No.:              5.97e-18        Length:         1972
Score:                  319.00          Matches:        144
Percent Similarity:     42.02%          Conservative:   114
Best Local Similarity:  23.45%          Mismatches:     180
Query Match:            12.83%          Indels:         176
DB:                     2               Gaps:           27

US-09-837-344-41 (1-1482) x US-08-875-435B-3 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        932 GluGluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA----------------------------------------------------AGCGAT 120
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
Db        992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG-------------------------- 336
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db       1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu---------------------- 1264

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db       1265 ----------------------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
```

```
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::|||      |||   :::   :::|||:::|||               :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr-----------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                :::  :::   |||||||||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                |||      :::::::|||:::          |||         |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                  ::::::           |||:::|||:::|||                |||   :::
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
                ::::::      :::   ::::::|||       |||   ||||||          ||||||
Db       1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------- 1290
                |||          |||   ||||||||||||:::
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1291 ------------------------------------AATAATTTTAAACCAAATGAT 1311
                                                 :::|||   :::
Db       1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT---------------------- 1350
                |||      :::|||:::   :::   :::   ||||||
Db       1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy       1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
                   |||:::      :::   :::||||||     |||:::|||      :::
Db       1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy       1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
                        |||   |||:::      :::||||||
Db       1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502

RESULT 18
US-09-538-092-918
; Sequence 918, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 918
;  LENGTH: 1937
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (0)...(0)
;  OTHER INFORMATION: Polypeptide Accession Number P13535
US-09-538-092-918

Alignment Scores:
Pred. No.:              7.88e-18        Length:         1937
Score:                  317.50          Matches:        156
Percent Similarity:     42.08%          Conservative:   131
Best Local Similarity:  22.87%          Mismatches:     184
Query Match:            12.77%          Indels:         211
DB:                     2               Gaps:           35

US-09-837-344-41 (1-1482) x US-09-538-092-918 (1-1937)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||::::::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
                |||:::|||::::::|||                         ::::::   |||
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                        |||   |||:::   :::   |||:::
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||:::|||||||||      |||||||||      ||||||:::
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                     :::|||||   |||
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||       |||            ||||||||||||:::
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||  ||||||                     ||||||
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
              |||:::|||   |||                   :::|||   |||::::::   :::
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ------------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                                           ||||||      :::::::::   ::::::
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   |||||          :::::::::|||         ::::::|||   ||||||
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA-----------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
              |||:::|||:::|||                  :::   :::|||      ::::::
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
                 :::   :::   |||:::   :::   |||||||||:::
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540

Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 -----------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                        :::|||   |||:::||||||:::::::::::|||:::|||:::
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                                    :::||||||      |||:::
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::      |||   ||||||           |||          |||:::   |||
Db       1250 LysMetCysArgSerLeuGlu-----------AspGlnValSerGluLeuLysThrLys 1265
```

```
Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
              :::|||            |||      ||||||              ||||||:::
Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
              |||   |||          :::   :::|||:::         :::          |||
Db       1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA------TCTATA 837
              |||   :::      :::   :::::::::      :::|||:::         :::
Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
              |||:::|||      :::   ::::::         ::::::      |||
Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
              |||:::   :::      ||||||:::::::::|||
Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy        933 ------------------------------------------------------------ 933

Db       1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy        934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                         :::   |||:::   :::   |||   |||
Db       1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
              |||   |||          :::          :::   |||   :::
Db       1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
              :::      |||   :::|||   ::::::|||   |||
Db       1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy       1087 ------------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                           |||:::   |||   |||:::|||
Db       1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------------- 1152
                 :::|||:::      |||   |||
Db       1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy       1153 ------------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176
                                      :::|||:::|||           ||||||
Db       1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy       1177 TTGATAGAGAAAAATGAAAATTTA-----------GATGATTTAGATGAAGGAATAGAA 1224
              |||      :::|||:::||||||              ||||||   |||   |||
Db       1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy       1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
              :::      :::::::::   |||   ||||||||||||||      :::   ||||||
Db       1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy       1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
              :::      :::      :::   ||||||      |||       |||
Db       1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy       1339 AAATATAAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
                    :::         ::::::|||:::   |||   :::      :::
Db       1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
              |||   :::   |||   ||||||           |||:::|||      :::
Db       1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy       1459 ACTAAA 1464
              |||:::
Db       1594 ThrArg 1595
```

```
RESULT 19
US-09-543-681A-6436
; Sequence 6436, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6436
;   LENGTH: 1180
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-6436

Alignment Scores:
Pred. No.:              9.75e-18      Length:         1180
Score:                  316.00        Matches:        98
Percent Similarity:     56.02%        Conservative:   51
Best Local Similarity:  36.84%        Mismatches:     77
Query Match:            12.71%        Indels:         40
DB:                     2             Gaps:           14

US-09-837-344-41 (1-1482) x US-09-543-681A-6436 (1-1180)

Qy          7 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTGCAAGAA 57
              :::|||          ||||||||||||   |||:::          |||:::   :::
Db        592 LysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeuGlnAlaGluGlnGluLysAla 611

Qy         58 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAAAAGTTACAAGAG 108
              :::|||          ||||||:::||||||||:::          |||:::   :::
Db        612 LysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeuGlnAlaGluGlnGluLysAla 631

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              :::|||          ||||||||||||||||||      :::||||||   :::|||
Db        632 LysGlnGluLysAlaGluGlnGluArgLeuAla------ArgLeuGlnAlaGluGlnGlu 649

Qy        169 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAG----------------------TTG 204
                    :::||||||:::      :::||||||                      |||
Db        650 LysAlaLysGlnGluLysAlaGluGlnGluLysAlaGluGlnGluArgLeuAlaArgLeu 669

Qy        205 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG--------- 255
              |||   :::|||          :::||||||:::      :::|||:::|||
Db        670 GlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeu 689

Qy        256 --------------CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                           :::   :::|||          ||||||:::|||||||||
Db        690 GlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAla------ 707

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::||||||   :::|||          :::||||||:::      :::|||:::|||
Db        708 ArgLeuGlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAla 727

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
                 |||:::      |||||||||:::      :::||||||   ::::::::::::
Db        728 ArgLeuGlnAla------GluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluLys 745

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              :::      ||||||||||||||||||      |||||||||   :::|||
Db        746 Ala------GluGlnGluArgLeuAla------LysLeuGlnAlaGluGlnGluLysAla 761

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||   |||      :::||||||   :::|||          |||||||||
Db        762 GluGlnGluArgLeuAla------ArgLeuGlnAlaGluGlnGluLysThrGluGlnGlu 779

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||   |||      :::||||||   :::|||          ||||||||||   |||
Db        780 ArgLeuAla------ArgLeuGlnAlaGluGlnGluLysAlaGluGlnGluArgLeuAla 797
```

```
Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
                 :::||||||   :::|||         ||||||   :::         :::
Db        798 ------ArgLeuGlnAlaGluGlnGluLysThrGluGlnGluArgLeu---AlaArgLeu 814

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                ||||||||||||         :::   ||||||      |||
Db        815 GlnAlaGluGlnGluLysAlaGluGlnGluArgLeuAlaArg---LeuGlnAlaAspLys 833

Qy        721 GAAATACCAGCTATAGAA 738
                |||:::|||   :::|||
Db        834 GluLeuProProValGlu 839

RESULT 20
US-10-144-198-2
; Sequence 2, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies Inc
;  TITLE OF INVENTION: Regulated Prostate Cance Genes
;  FILE REFERENCE: 9U 105 R1
;  CURRENT APPLICATION NUMBER: US/10/144,198
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1564
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-198-2

Alignment Scores:
Pred. No.:                1.11e-17      Length:          1564
Score:                    315.50        Matches:         115
Percent Similarity:       43.69%        Conservative:    110
Best Local Similarity:    22.33%        Mismatches:      163
Query Match:              12.69%        Indels:          127
DB:                       2             Gaps:            20

US-09-837-344-41 (1-1482) x US-10-144-198-2 (1-1564)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::::::::   :::   |||:::::::::|||             ::::::   |||
Db        629 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 648

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114
                   :::||||||:::|||         :::|||:::      ::::::
Db        649 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 668

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
                |||         ||||||   |||:::|||:::   :::|||::::::   |||
Db        669 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 687

Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
              |||:::||||||         |||   :::::::::   :::|||::::::   :::|||
Db        688 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 707

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||:::||||||   |||   :::|||:::   :::|||::::::   |||   |||:::
Db        708 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 727

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
              ||||||                  ::::::      |||   |||:::||||||
Db        728 AspArg-------------------AspArgAspHisAspArgGluArgGluArgGlu 740

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
                 :::|||:::   :::|||::::::   :::   |||:::|||   |||   :::|||
Db        741 ---ArgGluArgAspArgGluLysGluArgGluArgGluArgGluGluArgGluArgGlu 759

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::   :::|||::::::   :::   |||:::||||||
Db        760 ArgGluArgGluArgGluArgGluArgGluArgGluArg-------------------- 772
```

```
Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAA 519
                                  |||:::||||||   |||   :::::|||   :::|||:::
Db        773 -----------------GluArgGluArgAlaArgGluArgAspLysGluArgGluArg 786

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||   |||   |||   :::   :::   ::::::      :::|||::::::   :::
Db        787 GlnArgAspTrpGluAspLysAspLysGlyArgAspAspArgArgGluLysArgGluGlu 806

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              :::   ::::::|||      ::::::         |||::::::
Db        807 IleArgGluAspArgAsnProArgAspGlyHisAspGluArgLysSerLysLysArgTyr 826

Qy        640 AGGAAGGCTGATACG---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
              |||         :::         :::::::::   :::|||::::::||||||
Db        827 ArgAsnGluGlySerProSerProArgGlnSerProLysArgArgArgGluHisSer--- 845

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db        845 ------------------------------------------------------------ 845

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
                                                        |||      :::      |||
Db        846 ----------------------------------------ProAspSerAspAlaTyrAsn 852

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
              |||   |||   |||                |||||              |||
Db        853 SerGlyAspAspLysAsn------------GluLysHisArgLeuLeuSerGlnValVal 868

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--- 927
                   :::      |||   :::         ||||||   |||   :::   |||
Db        869 ArgProGlnGluSerArgSerLeuSerProSerHisLeuThrGluAspArgGlnGlyArg 888

Qy        928 ------------TCAATAAAACCAGAACAAAAAGAAGAT---------------------- 954
                          |||||||||:::||||||
Db        889 TrpLysGluGluAspArgLysProGluArgLysGluSerSerArgArgTyrGluGluGln 908

Qy        955 ------------------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
                               |||   |||   |||            |||   :::
Db        909 GluLeuLysGluLysValSerSerValAspLysGlnArgGluGlnThrGluIleLeuGlu 928

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
              |||   :::         |||   :::         :::                |||
Db        929 SerSerArgMetArgAlaGlnAspIleIleGlyHisHisGlnSerGluAspArgGluThr 948

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
              |||         ||||||   :::         :::   :::   |||||||||
Db        949 SerAspArgAlaHisAspGluAsnLysLysLysAlaLysIleGlnLys---LysProIle 967

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
                 :::         :::   |||:::::::::||||||   :::
Db        968 LysLys------------LysLysGluAspAspValGlyIleGluArg------------ 979

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
                                                        ||||||   :::||||||
Db        980 ----------------------------------GlyAsnIleGluThrThrSerGlu 987

Qy       1237 GAA---------TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
              :::         |||   :::|||   :::|||   |||       ||||||   :::
Db        988 AspGlyGlnValPheSerProLysLysGlyGlnLysLysLysSerIleGluLysLysArg 1007

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
                 :::         |||      :::   |||||         :::
Db       1008 LysLysSer------LysGlyAspSerAspIleSerAspGluGluAlaAlaGlnGlnSer 1025

Qy       1348 AATGATAAGCAG--------------------GTTAATAAGGAAAAGGAAAAATTCATA 1386
                   |||:::                    ||||||:::
Db       1026 LysLysLysArgGlyProArgThrProProIleThrThrLysGluGlu------------ 1041

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                 |||      :::   :::|||   |||   ||||||:::
Db       1042 ------LeuValGluMetCysAsnGlyLysAsnGlyIleLeuGlu 1054

RESULT 21
```

```
US-10-144-198-4
; Sequence 4, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies Inc
;  TITLE OF INVENTION: Regulated Prostate Cance Genes
;  FILE REFERENCE: 9U 105 R1
;  CURRENT APPLICATION NUMBER: US/10/144,198
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 1564
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-198-4

Alignment Scores:
Pred. No.:              1.11e-17        Length:         1564
Score:                  315.50          Matches:        115
Percent Similarity:     43.69%          Conservative:   110
Best Local Similarity:  22.33%          Mismatches:     163
Query Match:            12.69%          Indels:         127
DB:                     2               Gaps:           20

US-09-837-344-41 (1-1482) x US-10-144-198-4 (1-1564)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::::::::   :::   |||:::::::::|||            ::::::   |||
Db        629 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 648

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114
                   :::||||||:::|||         :::|||:::      ::::::
Db        649 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 668

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
                |||       ||||||   |||:::|||:::   :::|||::::::   |||
Db        669 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 687

Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
              |||:::||||||         |||   :::::::::   :::|||::::::   :::|||
Db        688 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 707

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||:::||||||   |||   :::|||:::   :::|||::::::   |||   |||:::
Db        708 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 727

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
              ||||||                  ::::::      |||   |||:::||||||
Db        728 AspArg--------------------AspArgAspHisAspArgGluArgGluArgGlu 740

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
                 :::|||:::   :::|||::::::   :::   |||:::|||   |||   :::|||
Db        741 ---ArgGluArgAspArgGluLysGluArgGluArgGluArgGluGluArgGluArgGlu 759

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::   :::|||::::::   :::   |||:::||||||
Db        760 ArgGluArgGluArgGluArgGluArgGluArgGluArg-------------------- 772

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAA 519
                          |||:::||||||   |||   :::::::|||   :::|||:::
Db        773 ------------------GluArgGluArgAlaArgGluArgAspLysGluArgGluArg 786

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||   |||   |||   :::   :::   ::::::      :::|||::::::   :::
Db        787 GlnArgAspTrpGluAspLysAspLysGlyArgAspAspArgArgGluLysArgGluGlu 806

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              :::   ::::::|||       ::::::          |||::::::
Db        807 IleArgGluAspArgAsnProArgAspGlyHisAspGluArgLysSerLysLysArgTyr 826

Qy        640 AGGAAGGCTGATACG---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
              |||         :::         :::::::::   :::|||::::::||||||
```

```
Db        827 ArgAsnGluGlySerProSerProArgGlnSerProLysArgArgArgGluHisSer--- 845

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db        845 ------------------------------------------------------------ 845

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
                                                  |||       :::      |||
Db        846 ------------------------------------ProAspSerAspAlaTyrAsn 852

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
              |||   |||   |||               ||||||           |||
Db        853 SerGlyAspAspLysAsn------------GluLysHisArgLeuLeuSerGlnValVal 868

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--- 927
                  :::      |||   :::        ||||||   |||   :::   |||
Db        869 ArgProGlnGluSerArgSerLeuSerProSerHisLeuThrGluAspArgGlnGlyArg 888

Qy        928 ------------TCAATAAAACCAGAACAAAAAGAAGAT---------------------- 954
                          |||||||||:::||||||
Db        889 TrpLysGluGluAspArgLysProGluArgLysGluSerSerArgArgTyrGluGluGln 908

Qy        955 ------------------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
                               |||   |||   |||             |||   :::
Db        909 GluLeuLysGluLysValSerSerValAspLysGlnArgGluGlnThrGluIleLeuGlu 928

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
                |||   :::           |||      :::                   |||
Db        929 SerSerArgMetArgAlaGlnAspIleIleGlyHisHisGlnSerGluAspArgGluThr 948

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
                |||          ||||||   :::            :::   :::   |||||||||
Db        949 SerAspArgAlaHisAspGluAsnLysLysLysAlaLysIleGlnLys---LysProIle 967

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
                :::           :::   |||:::::::::||||||   :::
Db        968 LysLys-----------LysLysGluAspAspValGlyIleGluArg------------ 979

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
                                                      ||||||   :::||||||
Db        980 ----------------------------------GlyAsnIleGluThrThrSerGlu 987

Qy       1237 GAA---------TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                :::            |||   :::|||   :::|||   |||       ||||||   :::
Db        988 AspGlyGlnValPheSerProLysLysGlyGlnLysLysLysSerIleGluLysLysArg 1007

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
                  :::              |||      :::   ||||||         :::
Db       1008 LysLysSer------LysGlyAspSerAspIleSerAspGluGluAlaAlaGlnGlnSer 1025

Qy       1348 AATGATAAGCAG--------------------GTTAATAAGGAAAAGGAAAAATTCATA 1386
                  |||:::                      ||||||:::
Db       1026 LysLysLysArgGlyProArgThrProProIleThrThrLysGluGlu------------ 1041

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                  |||       :::   :::|||   |||   ||||||:::
Db       1042 ------LeuValGluMetCysAsnGlyLysAsnGlyIleLeuGlu 1054

RESULT 22
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
```

```
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1077
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Alignment Scores:
Pred. No.:                1.26e-17        Length:          1960
Score:                    315.00          Matches:         145
Percent Similarity:       44.63%          Conservative:    100
Best Local Similarity:    26.41%          Mismatches:      168
Query Match:              12.67%          Indels:          136
DB:                       2               Gaps:            28

US-09-837-344-41 (1-1482) x US-09-538-092-1077 (1-1960)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::     ||||||:::|||     |||:::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::         |||             ||||||   :::||||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
                              |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::||||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                 |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
               :::||||||      |||::::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||      :::   :::|||:::||||||          :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                                |||     :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                  ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||          |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                  ||||||||||||||     :::
Db       1221 --------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
              ::::::      :::             :::|||:::|||         |||
Db       1231 AsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATATCTATAATAGAA--------------------------------AAAACAAAT 852
              |||      :::  |||                                 :::|||
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                           |||  ::::::          |||          |||   |||
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG------------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
              :::                          ||||||   |||             |||
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              :::|||             :::|||          :::   :::::::::|||
Db       1311 GlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                    ::::::   :::   ||||||||   ::::::   :::   :::   |||
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::|||||||||         :::|||:::      |||   |||
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                          :::         :::   |||::::::||||||   :::
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
                 |||      :::::::::         ::::::|||      |||   ||||||
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                    |||||||||:::            |||   ||||||||||||
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      |||   |||   :::|||      :::   ::::::
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||:::::::::||||||
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446

RESULT 23
US-09-949-016-10872
; Sequence 10872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
```

```
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10872
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10872

Alignment Scores:
Pred. No.:                1.26e-17      Length:          1960
Score:                    315.00        Matches:         145
Percent Similarity:       44.63%        Conservative:    100
Best Local Similarity:    26.41%        Mismatches:      168
Query Match:              12.67%        Indels:          136
DB:                       2             Gaps:            28

US-09-837-344-41 (1-1482) x US-09-949-016-10872 (1-1960)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||:::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::         |||            ||||||   :::||||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
                              |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::|||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                 |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                ::::||||||      |||:::::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||       :::   :::|||:::||||||          :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG---------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                                 |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||       |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      ||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                     ||||||   |||
```

```
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||             |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln---------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                 ||||||||||||||     :::
Db       1221 --------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
              ::::::      :::           :::|||:::|||         |||
Db       1231 AsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATATCTATAATAGAA---------------------------------AAAACAAAT 852
              |||      :::   |||                                 :::|||
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                        |||   ::::::      |||      |||   |||
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG----------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
              :::                        ||||||   |||            |||
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              :::|||            :::|||          :::   :::::::::|||
Db       1311 GlnAspThrGlnGluLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                    ::::::   :::   ||||||||   ::::::   :::   :::   |||
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::|||||||||          :::|||:::      |||   |||
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                          :::         :::   |||::::::||||||   :::
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
                |||      :::::::::          ::::::|||      |||   ||||||
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                    |||||||||:::           |||   ||||||||||||
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      |||   |||   ::::|||      :::   ::::::
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||::::::::::||||||
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446

RESULT 24
US-09-538-092-1078
; Sequence 1078, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
```

```
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1078
;   LENGTH: 1976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Alignment Scores:
Pred. No.:              1.52e-17        Length:         1976
Score:                  314.00          Matches:        146
Percent Similarity:     40.64%          Conservative:   108
Best Local Similarity:  23.36%          Mismatches:     179
Query Match:            12.63%          Indels:         192
DB:                     2               Gaps:           26

US-09-837-344-41 (1-1482) x US-09-538-092-1078 (1-1976)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||   |||   |||   |||   |||:::   ::::::   |||   ::::::|||:::
Db       1122 GluLeuGlnGluAspPheGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLys 1141

Qy         64 AGCGATTTAGAACAAGATAGACTTGCT---AAAGAAAAGTTACAAGAG------------ 108
                ||||||   ::::::      |||   |||   :::|||::::::
Db       1142 ArgAspLeuSerGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThr 1161

Qy        109 ------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 162
                   |||   :::|||      :::|||      ::::::::::::   |||
Db       1162 ThrAlaAlaGlnGlnGluLeuArgThrLysArg------GluGlnGluValAlaGluLeu 1179

Qy        163 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 198
              :::      ||||||:::|||   :::      |||
Db       1180 LysLysAlaLeuGluGluGluThrLysAsnHisGluAlaGlnIleGlnAspMetArgGln 1199

Qy        199 ------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 246
                         |||:::|||          |||||||||   :::|||   |||
Db       1200 ArgHisAlaThrAlaLeuGluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLys 1219

Qy        247 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT------AGACTTGCTAAAGAA 300
                |||::::::   :::      ||||||   |||         ||||||   |||
Db       1220 AlaAsnLeuGluLysAsnLysGlnGlyLeuGluThrAspAsnLysGluLeuAlaCysGlu 1239

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA--------- 351
                  ||||||          |||   |||   :::   |||:::
Db       1240 ValLysValLeuGlnGlnValLysAlaGluSerGluHisLysArgLysLysLeuAspAla 1259

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
              ::::::||||||      :::   :::   :::      |||   :::   :::|||
Db       1260 GlnValGlnGluLeuHisAlaLysValSerGluGlyAspArgLeuArgValGluLeuAla 1279

Qy        412 GAACAACAAAGCGATTTAGAACAAGAG------------------------------- 438
              |||:::   |||   |||:::   |||
Db       1280 GluLysAlaSerLysLeuGlnAsnGluLeuAspAsnValSerThrLeuLeuGluGluAla 1299

Qy        439 ---------------AGACTTGCTAAAGAA-----------------AAGTTGCAAGAA 465
                           :::   ||||||:::                  :::||||||:::
Db       1300 GluLysLysGlyIleLysPheAlaLysAspAlaAlaSerLeuGluSerGlnLeuGlnAsp 1319

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
                |||      |||::::::|||   |||   |||      |||      :::
Db       1320 ThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnLeuSerSerArgIleArg 1339

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
              ||||||:::|||          |||      ||||||||||||||          |||
Db       1340 GlnLeuGluGluGlu---------LysAsnSerLeuGlnGluGlnGln---------Glu 1353

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
              :::|||      |||::::::   |||::::::|||          |||:::   :::
Db       1354 GluGluGluGluAlaArgLysAsnLeuGluLysGlnValLeuAlaLeuGlnSerGlnLeu 1373

Qy        646 GCTGATACGAAAAAAAAATTTAGAA---------------------------------- 669
              ||||||||||||||   ::::::
Db       1374 AlaAspThrLysLysLysValAspAspAspLeuGlyThrIleGluSerLeuGluGluAla 1393

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              :::||||||         |||      ||||||   |||      ||||||||
Db       1394 LysLysLysLeuLeuLysAsp------AlaGluAlaLeuSerGlnArgLeuGlu------ 1409

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                                :::   :::|||
Db       1410 ---------------------------------------------GluLysAlaLeu 1413

Qy        790 CCTCAGGAC------AACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT------ATA 837
                  |||         :::      :::|||      :::|||:::
Db       1414 AlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuThr 1433

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA------------ 885
              ::::::     ::::::   :::  :::|||:::|||   :::
Db       1434 ValAspLeuAspHisGlnArgGlnValAlaSerAsnLeuGluLysLysGlnLysLysPhe 1453

Qy        886 --------------------AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                                  :::   |||         :::   ||||||:::
Db       1454 AspGlnLeuLeuAlaGluGluLysSerIleSerAlaArgTyrAlaGluGluArgAspArg 1473

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA--------------- 969
                  :::      :::||||||   |||:::      :::
Db       1474 AlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGlu 1493

Qy        970 --------------------------------CAAAATCATACATTAGAGACAGTAAAT 996
                                               ||||||    |||
Db       1494 GluAlaLeuGluAlaLysGluGluPheGluArgGlnAsnLysGlnLeuArg--------- 1510

Qy        997 ATTTCTGATGTTAATGATTTT------------------------------------- 1017
                 :::|||:::   |||
Db       1511 ---AlaAspMetGluAspLeuMetSerSerLysAspAspValGlyLysAsnValHisGlu 1529

Qy       1018 ------------------------------------------------CAAATAAGT 1026
                                                             |||:::
Db       1530 LeuGluLysSerLysArgAlaLeuGluGlnGlnValGluGluMetArgThrGlnLeuGlu 1549

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              :::   |||||||||:::   |||      |||      |||      |||
Db       1550 GluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMet 1569

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                           :::|||:::      :::|||      :::   |||:::||||||
Db       1570 GlnAlaMetLysAlaGlnPheGluArgAspLeuGlnThrArgAspGluGlnAsnGluGlu 1589

Qy       1147 AAC------------ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
                           |||         :::|||||||||      :::|||
Db       1590 LysLysArgLeuLeuIleLysGlnValArgGluLeuGluAlaGluLeuGluAspGluArg 1609

Qy       1195 ------------------------------------------------AATTTAGATGATTTA 1209
                                                             :::|||   ||||||
Db       1610 LysGlnArgAlaLeuAlaValAlaSerLysLysLysMetGluIleAspLeuLysAspLeu 1629

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
              :::      ||||||   ::::::   :::      :::|||   ||||||:::   :::
Db       1630 GluAlaGlnIleGlu---AlaAlaAsnLysAlaArgAspGluValIleLysGlnLeuArg 1648

Qy       1270 AAATATGAA---AAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
              |||   :::   :::   ||||||                                |||
Db       1649 LysLeuGlnAlaGlnMetLysAsp------------------------------TyrGln 1658

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAG------------GTTAATAAGGAAAAG 1374
                    :::::::::   :::      :::                :::||||||
Db       1659 ArgGluLeuGluGluAlaArgAlaSerArgAspGluIlePheAlaGlnSerLysGluSer 1678
```

```
Qy       1375 GAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATC 1434
Db       1679 GluLysLysLeuLysSerLeu-----------------GluAlaGluIleLeuGlnLeu 1692

Qy       1435 GTGGATGAGTTATCT 1449
Db       1693 GlnGluGluLeuAla 1697

RESULT 25
US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
;  APPLICANT: Hasegawa, Kazuhide
;  APPLICANT:  Arakawa, Emi
;  APPLICANT:  Oda, Shoji
;  APPLICANT:  Matsuda, Yuzuru
;  APPLICANT:  Takahashi, Katsuhito
;  APPLICANT:  Sugahara, Michihiro
;  APPLICANT:  Ishiyama, Haruo
;  TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
;  TITLE OF INVENTION:  FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
;  TITLE OF INVENTION:  VECTOR DNA, MICROORGANSIM CARRYING THE RECOMBINANT DNA, AND
;  TITLE OF INVENTION:  AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
;  TITLE OF INVENTION:  RECOMBINANT DNA
;  FILE REFERENCE: 07898-013001
;  CURRENT APPLICATION NUMBER: US/08/875,435B
;  CURRENT FILING DATE:  1997-07-25
;  PRIOR APPLICATION NUMBER: PCT/JP96/00134
;  PRIOR FILING DATE: 1996-01-25
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Alignment Scores:
Pred. No.:              3.86e-17      Length:          1972
Score:                  309.00        Matches:         148
Percent Similarity:     41.96%        Conservative:    113
Best Local Similarity:  23.79%        Mismatches:      169
Query Match:            12.43%        Indels:          192
DB:                     2             Gaps:            29

US-09-837-344-41 (1-1482) x US-08-875-435B-4 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-----------------------------------------------AGCGAT 120
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAAGAA---------------------- 198
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        199 ----------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAA 228
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        229 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT--- 285

Db       1052 GluLeuGluLysLeuLysArgLysMetAspGlyGluAlaSerAspLeuHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAA-----------CAAGAGAGACGTGCTAAAGAA 351
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
Db       1152 LysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAA 570
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnValValGluGluLeuThrGluGln 1211

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
Db       1212 ---------LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThr 1228

Qy        631 TTAGAACAAAGGAAGGCTGAT-----------------------------ACGAAAAAA 660
Db       1229 LeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1248

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db       1249 GluValGluHisLysLysLys-------------------------------LysLeu 1257

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db       1258 GluValGlnLeuGlnGluLeu-----------------------------------Gln 1265

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
Db       1266 SerLysCysSerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeu 1285

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db       1286 Gln---AsnGluValGluSerValThr---------------------------------- 1293

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA----------------------- 936
Db       1294 GlyMetLeuSerGluAlaGluGlyLysAlaIleLysLeuAlaLysGluValAlaSerLeu 1313

Qy        937 CCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
Db       1314 GlySerGlnLeuGlnAspThrGlnGluLeuGlnGluGluThrArgGlnLysLeuAsn 1333

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
Db       1334 ValSerThr-----------LysLeuArgGlnLeuGluAspGlu--------------- 1344

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
Db       1345 ArgAsnSerLeu------GlnGluGlnLeuAspGluGluMetGluAlaLysGlnAsnLeu 1362

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
```

```
Db       1363 GluArgHisIleSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAsp 1382

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236

Db       1383 PheAlaSerThrValGluSerLeuGluGluGlyLysLysArgPheGlnLysGluIleGlu 1402

Qy       1237 GAATTATCT---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287

Db       1403 SerLeuThrGlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLys 1422

Qy       1288 GAT------------------------------------------------------------ 1290

Db       1423 AsnArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGlnArgGlnLeu 1442

Qy       1291 ---AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347

Db       1443 ValSerAsnLeuGluLysLysGlnLysLys---PheAspGlnLeuLeuAlaGluGluLys 1461

Qy       1348 AAT-------------------------GATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Db       1462 AsnIleSerSerLysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLys 1481

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Db       1482 GluThrLysAlaLeu---SerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGlu 1500

Qy       1441 GAGTTA 1446

Db       1501 GluLeu 1502
```

RESULT 26
US-09-949-016-8135
; Sequence 8135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8135
; LENGTH: 1942
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8135

Alignment Scores:

Pred. No.:	6.13e-17	Length:	1942
Score:	306.50	Matches:	154
Percent Similarity:	41.06%	Conservative:	140
Best Local Similarity:	21.51%	Mismatches:	185
Query Match:	12.33%	Indels:	237
DB:	2	Gaps:	37

US-09-837-344-41 (1-1482) x US-09-949-016-8135 (1-1942)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60

Db        934 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 953

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93

Db        954 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 973

Qy         94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126

Db        974 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 993

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168

Db        994 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1013

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201

Db       1014 LeuAspAspLeuGlnMetGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1033

Qy        202 TTGCAAGAACAACAAAGCGAT-----------TTAGAACAAGAGAGA------------ 237

Db       1034 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuCysMet 1053

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267

Db       1054 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1073

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309

Db       1074 MetAspThrGluAsnAspLysGlnGlnLeuAsnGluLysLeuLysLysLysGluPheGlu 1093

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339

Db       1094 MetSerAsnLeuGlnGlyLysIleGluAspGluGlnAlaLeuAlaIleGlnLeuGlnLys 1113

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390

Db       1114 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1133

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447

Db       1134 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1153

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495

Db       1154 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1173

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------------- 540

Db       1174 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluSerThrLeuGlnHisGlu 1193

Qy        540 ------------------------------------------------------------ 540

Db       1194 AlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1213

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588

Db       1214 IleAspSerLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1233

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615

Db       1234 GluIleAsnAspLeuAlaSerAsnMetGluThrValSerLysAlaLysAlaAsnPheGlu 1253

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTAGAAAGA 672

Db       1254 LysMetCysArgThrLeuGluAspGlnLeuSerGluIleLysThrLysGluGluGluGln 1273

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732

Db       1274 GlnArgLeuIleAsnGluLeuSerAlaGlnLys------AlaArgLeuHisThrGluSer 1291

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792

Db       1292 GlyGluPheSerArgGlnLeuAspGlu--------------LysAspAlaMetValSer 1306

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA--------------TCTATA 837

Db       1307 GlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLysArgGlnLeu 1326

Qy        838 ATAGAAAAAACAAAT---------------------------------------------- 852
```

```
                 |||:::|||
Db       1327 GluGluGluThrLysAlaLysSerThrLeuAlaHisAlaLeuGlnSerAlaArgHisAsp 1346

Qy        853 ------------AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                          ||||||                  |||   :::   ::::::   :::
Db       1347 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluAlaLysAlaGluLeuGlnArg 1366

Qy        901 GGA--------------------------------CATCTTGAAGAAAAGAAAGATGGT 927
              |||                                    :::   :::   |||
Db       1367 GlyMetSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1386

Qy        928 TCAATAAAACCAGAACAAAAAGAAGAT--------------------AAATCTGCTGAC 966
                  :::   |||:::   |||:::                      :::     |||
Db       1387 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAsp 1406

Qy        967 ATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGT 1026
                 :::   |||   :::|||   ||||||           :::          :::
Db       1407 AlaGluGluHis---ValGluAlaValAsnSerLysCysAlaSer---------LeuGlu 1422

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              |||   :::       :::       |||   :::|||   :::||||||   |||
Db       1423 LysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGluArgSer 1442

Qy       1087 ----------------------------------------GATGAAGACTTAGACGAA 1104
                                                     |||:::   |||   |||
Db       1443 AsnAlaAlaCysIleAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuAlaGlu 1462

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA------------ 1152
              :::|||          :::|||:::       |||   |||
Db       1463 TrpLysGln------LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGlu 1480

Qy       1153 ----------------------------------GGAATTTATAAAGAA--------- 1167
                                                 |||:::|||
Db       1481 SerArgSerLeuSerThrGluLeuPheLysValLysAsnAlaTyrGluGluSerLeuAsp 1500

Qy       1168 ---CTAGAAGATTTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGAT 1212
                 ||||||   |||      :::|||:::||||||              ||||||
Db       1501 HisLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThr 1520

Qy       1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAG 1269
              |||   |||   :::      :::   :::   |||   ||||||:::||||||
Db       1521 GluGlnIleAlaGluGlyGlyLysHisIleHisGluLeuGluLysValLysLysGlnLeu 1540

Qy       1270 AAATATGAAAAAACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
                 :::||||||::::::         :::   :::      :::   ||||||
Db       1541 AspHisGluLysSerGluLeuGlnThrSerLeuGluGluAlaGluAlaSerLeuGluHis 1560

Qy       1327 GAGCATATTAAA-------------------------------AAATATAAAAATGAT 1353
              |||      |||                                |||   :::   |||
Db       1561 GluGluGlyLysIleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluIleAsp 1580

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAA----------------------------AAATTC 1383
              :::::::::   ::::::   |||                            :::
Db       1581 ArgLysIleAlaGluLysAspGluGluLeuAspGlnLeuLysArgAsnHisLeuArgVal 1600

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGAC------------AATGAAATTTTACAG 1431
              ::::::|||:::            |||   :::            |||:::   |||:::
Db       1601 ValGluSerMetGlnSerThrLeuAspAlaGluIleArgSerArgAsnAspAlaLeuArg 1620

Qy       1432 ATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||      ::::::      |||:::   :::      :::::::|||
Db       1621 IleLysLysLysMetGluGlyAspLeuAsnGluMetGluIleGlnLeu 1636

RESULT 27
US-09-538-092-915
; Sequence 915, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 915
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P12882
US-09-538-092-915

Alignment Scores:
Pred. No.:              8.11e-17      Length:          1939
Score:                  305.00        Matches:         159
Percent Similarity:     41.53%        Conservative:    123
Best Local Similarity:  23.42%        Mismatches:      163
Query Match:            12.27%        Indels:          234
DB:                     2             Gaps:            39

US-09-837-344-41 (1-1482) x US-09-538-092-915 (1-1939)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        931 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 950

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                 |||:::|||::::::|||                            ::::::   |||
Db        951 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 970

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                        |||   |||:::   :::   |||:::
Db        971 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 990

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        991 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1010

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                          :::||||||   |||
Db       1011 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysIleLys 1030

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||            |||||||||||:::
Db       1031 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysIleArgMet 1050

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||   ||||||                    ||||||
Db       1051 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerAla 1070

Qy        268 AGCGATTTAGAACAAGATAGA---CTTGCTAAAGAAAAGTTACAAGAGCAG--------- 315
                 |||:::|||   |||:::           |||||||||:::::::::
Db       1071 MetAspIleGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysGluPheGlu 1090

Qy        316 ------------CAAAGCGATTTAGAACAAGAG-----------------------AGA 339
                         ||||||   :::|||   |||                       :::
Db       1091 MetSerGlyLeuGlnSerLysIleGluAspGluGlnAlaLeuGlyMetGlnLeuGlnLys 1110

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||         ::::::::|||   :::   ::::::|||   ||||||
Db       1111 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1130

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1131 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1150

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
                 |||:::|||:::|||                 :::   :::|||      ::::::
```

```
Db       1151 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1170

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---------------- 540
              :::   :::   |||:::   :::   |||||||||:::
Db       1171 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1190

Qy        540 ------------------------------------------------------------ 540

Db       1191 AlaThrAlaAlaThrLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1210

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                         :::|||   |||:::||||||:::::::::::|||:::::::::
Db       1211 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluMetLysMet 1230

Qy        589 GAG---------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                                 :::|||||||      |||:::
Db       1231 GluIleAspAspLeuAlaSerAsnMetGluThrValSerLysAlaLysGlyAsnLeuGlu 1250

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTAGAAAGA 672
              :::      |||   ||||||   :::   ::::::   |||   |||      |||:::
Db       1251 LysMetCysArgAlaLeuGluAspGlnLeuSerGluIleLysThrLysGluGluGluGln 1270

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ::::::          |||:::   |||:::                |||||
Db       1271 GlnArgLeuIleAsnAspLeuThrAlaGln------ArgAlaArgLeu------------ 1284

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTAT---------ATACCACATCAATCT 783
                        :::   |||   |||   |||         :::      :::
Db       1285 ---------------GlnThrGluSerGlyGluTyrSerArgGlnLeuAspGluLysAsp 1299

Qy        784 TCTTTA---CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              :::|||      |||   :::||||||           :::::::::|||      :::
Db       1300 ThrLeuValSerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeu 1319

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ::::::          |||   |||          :::   :::      :::
Db       1320 LysArgGlnLeuGluGluGluIle----------LysAlaLysSerAlaLeuAlaHis 1335

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA--- 957
                        :::   |||       :::       ||||||   |||::::::
Db       1336 AlaLeuGlnSerSerArgHisAspCysAspLeuLeuArgGluGlnTyrGluGluGluGln 1355

Qy        958 ------TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAAT 1011
                  |||::::::|||
Db       1356 GluAlaLysAlaGluLeuGlnArg---------------------------------- 1363

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT---------GCTGAATATGACGAT 1062
                 :::||||||          |||::::::              :::|||:::
Db       1364 ------AlaMetSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThr 1381

Qy       1063 TCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA 1122
                |||         |||:::   |||:::             |||   :::   |||
Db       1382 AspAlaIleGlnArgThrGluGluLeuGluGluAlaLys------LysLysLeuAlaGln 1399

Qy       1123 TATGACAATTTCCAAGAT---GAAGAAAACATA------------------------- 1152
                         ||||||   ||||||::::::
Db       1400 ------ArgLeuGlnAspAlaGluGluHisValGluAlaValAsnAlaLysCysAlaSer 1417

Qy       1153 ------------------GGAATTTATAAAGAACTAGAAGATTTGATA------------ 1182
                               :::         |||:::||||||||||:::
Db       1418 LeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGlu 1437

Qy       1183 ----------------------------GAGAAAAATGAAAATTTAGATGATTTA--- 1209
                                         :::|||      |||   |||   :::
Db       1438 ArgThrAsnAlaAlaCysAlaAlaLeuAspLysLysGlnArgAsnPheAspLysIleLeu 1457

Qy       1210 ------------------------------------GATGAAGGAATAGAAAAATCA 1230
                                                    |||      :::|||
Db       1458 AlaGluTrpLysGlnLysCysGluGluThrHisAlaGluLeuGluAlaSerGlnLysGlu 1477

Qy       1231 TCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT 1290
              |||      ||||||   |||      |||      ||||||::::::   |||
Db       1478 SerArgSerLeuSerThrGluLeuPheLysIleLysAsnAlaTyrGluGluSerLeuAsp 1497
```

```
Qy       1291 ------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG--------------- 1329
                         |||      :::|||:::|||      |||
Db       1498 GlnLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThr 1517

Qy       1330 --------------------------------CATATTAAAAAATATAAAAATGATAAG 1356
                                              ::::::|||   |||      |||
Db       1518 GluGlnIleAlaGluGlyGlyLysArgIleHisGluLeuGluLysIleLys------Lys 1535

Qy       1357 CAGGTTAATAAGGAAAAG--------------------GAAAAATTCATAAAATCATTG 1395
                |||||   :::||||||                   |||:::         ||||||
Db       1536 GlnValGluGlnGluLysSerGluLeuGlnAlaAlaLeuGluGluAlaGluAlaSerLeu 1555

Qy       1396 TTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAA 1452
                 |||       :::   :::||||||:::|||      ||||||::::::
Db       1556 GluHis---------GluGluGlyLysIleLeuArgIleGlnLeuGluLeuAsnGln 1571

RESULT 28
US-09-949-016-11104
; Sequence 11104, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11104
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-11104

Alignment Scores:
Pred. No.:                8.11e-17        Length:          1939
Score:                    305.00          Matches:         159
Percent Similarity:       41.53%          Conservative:    123
Best Local Similarity:    23.42%          Mismatches:      163
Query Match:              12.27%          Indels:          234
DB:                       2               Gaps:            39

US-09-837-344-41 (1-1482) x US-09-949-016-11104 (1-1939)

Qy           4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
                 |||   :::   ::::::       |||   ·   ||||||:::   ||||||:::::::::
Db         931 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 950

Qy          61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
                  |||:::|||::::::|||                                 ::::::   |||
Db         951 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 970

Qy          94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
                  ||||||                                |||   |||:::   :::   |||:::
Db         971 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 990

Qy         127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
                  :::          :::|||   |||||||||      |||||||||      ||||||:::
Db         991 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1010

Qy         169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                          :::||||||   |||
Db        1011 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysIleLys 1030

Qy         202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
```

```
Db       1031 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysIleArgMet 1050

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1051 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1070

Qy        268 AGCGATTTAGAACAAGATAGA---CTTGCTAAAGAAAAGTTACAAGAGCAG--------- 315
Db       1071 MetAspIleGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysGluPheGlu 1090

Qy        316 ------------CAAAGCGATTTAGAACAAGAG-----------------------AGA 339
Db       1091 MetSerGlyLeuGlnSerLysIleGluAspGluGlnAlaLeuGlyMetGlnLeuGlnLys 1110

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1111 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1130

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db       1131 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1150

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1151 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1170

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------------- 540
Db       1171 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1190

Qy        540 ------------------------------------------------------------ 540
Db       1191 AlaThrAlaAlaThrLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1210

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       1211 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluMetLysMet 1230

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
Db       1231 GluIleAspAspLeuAlaSerAsnMetGluThrValSerLysAlaLysGlyAsnLeuGlu 1250

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTAGAAAGA 672
Db       1251 LysMetCysArgAlaLeuGluAspGlnLeuSerGluIleLysThrLysGluGluGluGln 1270

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
Db       1271 GlnArgLeuIleAsnAspLeuThrAlaGln------ArgAlaArgLeu------------ 1284

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTAT---------ATACCACATCAATCT 783
Db       1285 --------------GlnThrGluSerGlyGluTyrSerArgGlnLeuAspGluLysAsp 1299

Qy        784 TCTTTA---CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
Db       1300 ThrLeuValSerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeu 1319

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db       1320 LysArgGlnLeuGluGluGluIle------------LysAlaLysSerAlaLeuAlaHis 1335

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA--- 957
Db       1336 AlaLeuGlnSerSerArgHisAspCysAspLeuLeuArgGluGlnTyrGluGluGluGln 1355

Qy        958 ------TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAAT 1011
Db       1356 GluAlaLysAlaGluLeuGlnArg---------------------------------- 1363

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT---------GCTGAATATGACGAT 1062
Db       1364 ------AlaMetSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThr 1381

Qy       1063 TCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA 1122
Db       1382 AspAlaIleGlnArgThrGluGluLeuGluGluAlaLys------LysLysLeuAlaGln 1399

Qy       1123 TATGACAATTTCCAAGAT---GAAGAAAACATA-------------------------- 1152
Db       1400 ------ArgLeuGlnAspAlaGluGluHisValGluAlaValAsnAlaLysCysAlaSer 1417

Qy       1153 -----------------GGAATTTATAAAGAACTAGAAGATTTGATA------------ 1182
Db       1418 LeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGlu 1437

Qy       1183 ----------------------------GAGAAAAATGAAAATTTAGATGATTTA--- 1209
Db       1438 ArgThrAsnAlaAlaCysAlaAlaLeuAspLysLysGlnArgAsnPheAspLysIleLeu 1457

Qy       1210 -------------------------------------GATGAAGGAATAGAAAAATCA 1230
Db       1458 AlaGluTrpLysGlnLysCysGluGluThrHisAlaGluLeuGluAlaSerGlnLysGlu 1477

Qy       1231 TCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT 1290
Db       1478 SerArgSerLeuSerThrGluLeuPheLysIleLysAsnAlaTyrGluGluSerLeuAsp 1497

Qy       1291 ------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG--------------- 1329
Db       1498 GlnLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThr 1517

Qy       1330 --------------------------------CATATTAAAAAATATAAAAATGATAAG 1356
Db       1518 GluGlnIleAlaGluGlyGlyLysArgIleHisGluLeuGluLysIleLys------Lys 1535

Qy       1357 CAGGTTAATAAGGAAAAG--------------------GAAAAATTCATAAAATCATTG 1395
Db       1536 GlnValGluGlnGluLysSerGluLeuGlnAlaAlaLeuGluGluAlaGluAlaSerLeu 1555

Qy       1396 TTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAA 1452
Db       1556 GluHis---------GluGluGlyLysIleLeuArgIleGlnLeuGluLeuAsnGln 1571
```

RESULT 29
US-09-949-016-6925
; Sequence 6925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6925
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6925

Alignment Scores:

Pred. No.:	8.91e-17	Length:	1939
Score:	304.50	Matches:	154
Percent Similarity:	41.06%	Conservative:	140
Best Local Similarity:	21.51%	Mismatches:	185
Query Match:	12.25%	Indels:	237

```
DB:                 2          Gaps:         37

US-09-837-344-41 (1-1482) x US-09-949-016-6925 (1-1939)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        931 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 950

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                |||:::|||::::::|||                          ::::::   |||
Db        951 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 970

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                        |||   |||:::   :::   |||:::
Db        971 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 990

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        991 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1010

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                    :::||||||   |||
Db       1011 LeuAspAspLeuGlnMetGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1030

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||            ||||||||||||:::
Db       1031 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuCysMet 1050

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||   ||||||                    ||||||
Db       1051 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1070

Qy        268 AGCGATTTAGAACAAGAT------------------AGACTTGCTAAAGAAAAGTTACAA 309
                 |||   |||   |||                 :::|||   |||::::::   :::
Db       1071 MetAspThrGluAsnAspLysGlnGlnLeuAsnGluLysLeuLysLysLysGluPheGlu 1090

Qy        310 ------------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                                           ||||||        ::::::   ::::::
Db       1091 MetSerAsnLeuGlnGlyLysIleGluAspGluGlnAlaLeuAlaMetGlnLeuGlnLys 1110

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||         :::::::::|||   :::   ::::::|||   ||||||
Db       1111 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1130

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1131 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1150

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
                 |||:::|||:::|||                :::   :::|||      ::::::
Db       1151 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluLeuAsnLysLys 1170

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------------- 540
                 :::   :::   |||:::   :::   |||||||||:::
Db       1171 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluSerThrLeuGlnHisGlu 1190

Qy        540 ------------------------------------------------------------ 540

Db       1191 AlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyLysGln 1210

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                         :::|||   |||:::||||||:::::::::::|||:::|||:::
Db       1211 IleAspSerLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1230

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                               :::||||||          :::
Db       1231 GluIleAsnAspLeuAlaSerAsnMetGluThrValSerLysAlaLysAlaAsnPheGlu 1250

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTAGAAAGA 672
              :::      |||   ||||||   :::   ::::::   |||   |||      |||:::
Db       1251 LysMetCysArgThrLeuGluAspGlnLeuSerGluIleLysThrLysGluGluGluGln 1270

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732

              ::::::         ::::::   |||:::            ||||||         :::
Db       1271 GlnArgLeuIleAsnGluLeuSerAlaGlnLys------AlaArgLeuHisThrGluSer 1288

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
                |||          :::   :::                              :::   :::
Db       1289 GlyGluPheSerArgGlnLeuAspGlu--------------LysAspAlaMetValSer 1303

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA--------------TCTATA 837
              |||   :::||||||          :::::::::|||                  :::
Db       1304 GlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLysArgGlnLeu 1323

Qy        838 ATAGAAAAAACAAAT---------------------------------------------- 852
                |||:::|||
Db       1324 GluGluGluThrLysAlaLysSerThrLeuAlaHisAlaLeuGlnSerAlaArgHisAsp 1343

Qy        853 ------------AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                        ||||||                 |||   :::   ::::::.  :::
Db       1344 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluAlaLysAlaGluLeuGlnArg 1363

Qy        901 GGA--------------------------------CATCTTGAAGAAAAGAAAGATGGT 927
              |||                                :::   :::   |||
Db       1364 GlyMetSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1383

Qy        928 TCAATAAAACCAGAACAAAAAGAAGAT--------------------AAATCTGCTGAC 966
                 :::   |||:::   |||:::                        :::       |||
Db       1384 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAsp 1403

Qy        967 ATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGT 1026
                 :::   |||   :::|||   ||||||          :::         :::
Db       1404 AlaGluGluHis---ValGluAlaValAsnSerLysCysAlaSer---------LeuGlu 1419

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              |||   :::       :::   |||   :::|||   :::|||||   |||
Db       1420 LysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGluArgSer 1439

Qy       1087 ----------------------------------------GATGAAGACTTAGACGAA 1104
                                                     |||:::   |||   |||
Db       1440 AsnAlaAlaCysIleAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuAlaGlu 1459

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA------------ 1152
              :::|||       :::|||:::       |||   |||
Db       1460 TrpLysGln------LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGlu 1477

Qy       1153 ------------------------------------GGAATTTATAAAGAA--------- 1167
                                                 |||:::|||
Db       1478 SerArgSerLeuSerThrGluLeuPheLysValLysAsnAlaTyrGluGluSerLeuAsp 1497

Qy       1168 ---CTAGAAGATTTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGAT 1212
                 ||||||   |||      :::|||:::||||||               ||||||
Db       1498 HisLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThr 1517

Qy       1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAG 1269
              |||   |||   :::      :::   :::   |||   ||||||:::||||||
Db       1518 GluGlnIleAlaGluGlyGlyLysHisIleHisGluLeuGluLysValLysLysGlnLeu 1537

Qy       1270 AAATATGAAAAAACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
                 :::||||||:::::         :::   :::      :::   ||||||
Db       1538 AspHisGluLysSerGluLeuGlnThrSerLeuGluGluAlaGluAlaSerLeuGluHis 1557

Qy       1327 GAGCATATTAAA--------------------------------AAATATAAAAATGAT 1353
              |||      |||                                |||   :::   |||
Db       1558 GluGluGlyLysIleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluIleAsp 1577

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAA----------------------------AAATTC 1383
              :::::::::   ::::::   |||                           :::
Db       1578 ArgLysIleAlaGluLysAspGluGluLeuAspGlnLeuLysArgAsnHisLeuArgVal 1597

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGAC------------AATGAAATTTTACAG 1431
              ::::::|||:::            |||   :::            |||:::   |||:::
Db       1598 ValGluSerMetGlnSerThrLeuAspAlaGluIleArgSerArgAsnAspAlaLeuArg 1617

Qy       1432 ATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||      ::::::      |||:::   :::      ::::::|||
```

```
Db       1618 IleLysLysLysMetGluGlyAspLeuAsnGluMetGluIleGlnLeu 1633

RESULT 30
US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
;   LENGTH: 924
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-18798

Alignment Scores:
Pred. No.:              9.71e-17        Length:         924
Score:                  303.50          Matches:        129
Percent Similarity:     43.24%          Conservative:   111
Best Local Similarity:  23.24%          Mismatches:     196
Query Match:            12.21%          Indels:         119
DB:                     2               Gaps:           21

US-09-837-344-41 (1-1482) x US-09-248-796A-18798 (1-924)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||:::         :::|||      |||      :::|||:::::::::::::|||:::
Db        325 GlnGlnPheThrLysGluLeuIleAlaGluSerGluSerLysGlnGlnValGluGluGlu 344

Qy         61 CAAAGCGATTTAGAA------------CAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                 :::|||:::           :::|||                 :::|||:::
Db        345 LeuGluAsnLeuLysThrIleHisAlaLysAspSerAlaArgIleLeuGluLeuGluThr 364

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              |||   ||||||      :::|||:::         :::                   |||
Db        365 GlnLeuSerAspAlaAlaLysGluLysSerGluSerAspTyrLysLeuThrAspThrSer 384

Qy        169 GATCTAGAACAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAACAACAAAGCGATTTA 225
              ::::::      :::   :::::::::   |||   |||:::          :::   |||
Db        385 GluIleValAsnAspLeuLysSerGlnIleGluThrLeuLysAlaAsnLeuAsnLysLeu 404

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAA 282
              |||:::||||||      :::   |||      :::|||   |||:::|||   |||
Db        405 GluGluGluArgGluIleGlnAsnLysLysLeuAspGlnValSerGluLeuLysGluLeu 424

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
                  |||      |||:::|||                 |||:::||||||
Db        425 LysGluLeuLysValGluGluLeuSerAsnAsnLeuLeuLysLeuGlnGlnGluLeuHis 444

Qy        343 GCTAAAGAA---AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 399
                ||||||   :::|||   ||||||          |||                   |||
Db        445 ArgLysGluIleGluLeuAsnGluGlnLeuGluLysLeuHisGlyValSerAlaAspLys 464

Qy        400 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 459
              |||||||||   :::|||         :::||||||   ::::::          |||
Db        465 GluLysLeuLeuGlnGln------SerIleGluGlnSerLysIleTyrIleAsnSerLeu 482

Qy        460 CAAGAACAACAAAGCGATTTAGAACAA--------------------------GAGAGA 492
              ::::::|||   :::      ::::::                          |||:::
Db        483 LysAspGlnAsnAlaLysThrAspArgAsnTrpGluIleLysTyrGlyThrLeuGluGln 502

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                  ::::::   :::::::|||      ::::::   ::::::::::::
Db        503 AsnGlnGlnLysLeuValGluGluHisLeuAlaAsnArgGlnArgAspGlnLeuGluIle 522

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAG 609
              ||||||||   :::   :::::::::||||||   :::   :::|||   |||   :::
Db        523 GluLysLeuValLysSerLysThrGluLeuGluAlaAspAsnGlnAlaLeuLysAsnGln 542

Qy        610 TTGCAAGAG---------CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||   |||         |||       |||              ::::::   ||||||
Db        543 LeuThrGluIleSerLysGlnPheSerThrLeuGlyValGluSerSerGluLeuLysLys 562

Qy        661 AATTTAGAAAGA---------AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711
              ||||||           |||   :::      :::|||:::   ::::::|||
Db        563 GlnLeuGluGluThrGlnIleLysTyrAspGlnAlaGluIleValHisGlnGluLeuLys 582

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771
              ::::::|||
Db        583 AlaGlnIleGlu---------------------------------------------- 586

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA 831
                        |||   :::::::::   |||             ||||||:::
Db        587 ------------LysLeuLeuLysGluSerAlaGlyLysAspGlyGlnLeuLysGluLeu 602

Qy        832 TCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGAT 891
              :::                     :::|||             |||         ::::::
Db        603 Asn--------------------GlnSerHisGluLysPheValAsnGluLeuLysAsn 615

Qy        892 ATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951
              |||:::      ::::::|||   ||||||      ||||||      :::::::::
Db        616 GluHisGluLysGlnValLysGluThrLysAspGlnIleIleLysGluMetGluGluLys 635

Qy        952 GATAAATCTGCT--------------GACATACAAAATCATACATTAGAGACAGTAAAT 996
                 :::   |||               |||      |||      ::::::   :::|||
Db        636 HisGlnGlnAlaIleLysGluIleGluAspSerHisAsnGluAsnIleLysGluIleAsn 655

Qy        997 ----------------------------ATTTCTGATGTTAATGAT------------ 1014
                                          |||      :::|||:::
Db        656 AsnGluHisGluAsnLysAlaLysCysIleIleAspSerLeuAsnGluGluIleGluGlu 675

Qy       1015 ------TTTCAAATAAGTAAGTATGAGGATGAA-----------------ATAAGTGCT 1050
                    |||:::         |||   |||                     :::
Db        676 LeuThrSerGlnLeuLysAsnAlaGluSerGluLysAsnThrLeuGlnSerLeuLysLeu 695

Qy       1051 GAATATGACGATTCATTAATAGAT-----------------------GAAGAAGAAGAT 1086
              ||||||::::::   :::|||                        |||:::|||
Db        696 GluTyrGluAsnGluIleIleAlaTyrLysSerLysIleAspGlnLeuGluLysGluSer 715

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                 |||:::|||   |||::::::      :::|||          |||       |||
Db        716 AlaGluAsnLeuLysGluTyrGluAlaLysLeuGlnSerMetLysPheAspLeuGluSer 735

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGAT---------------------------- 1176
              ::::::   |||   |||:::|||
Db        736 AspLeuAlaIleGluLysGlnLeuArgLysAspAspGlyGlnAspPheGluAsnGlnIle 755

Qy       1177 -----------------TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAA--- 1215
                              :::   ||||||:::      |||   :::   ||||||
Db        756 GluLysLeuAsnGlnLeuValThrGluLysAspLeuGlnLeuSerGluLysAspGluGlu 775

Qy       1216 ------GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
                    |||:::|||      |||||||||   :::   |||   |||      :::
Db        776 IleAlaSerIleLysLysTyrMetGluGluLeuGluAspThrLysAlaLysLeuGluGlu 795

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
              |||          |||      |||      |||      |||   ||||||      :::
Db        796 Lys------SerThrThrGlnAsnGlyMetLysLeuLeuAspAsnSerLeuThrLysLys 813

Qy       1330 CAT------ATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
              |||      :::           :::   |||      |||   :::   :::
Db        814 HisAlaAlaValValAlaGluLeuSerGlyLysIleThrAsnGlyAspAlaAsp------ 831

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
                                   |||||||||||||||   |||
Db        832 --------------------IleAspGlyAspAsnGluProLeu 839
```

Search completed: December 21, 2005, 09:37:12
Job time : 126.334 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:23 ; Search time 234.352 Seconds
(without alignments)
5284.546 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22847/app_query.fasta_1.3740
-DB=Published_Applications_AA_Main -QFMT=fastan -SUFFIX=rapbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_781_@runat_21122005_093214_22847
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2406.5	96.8	1909	5	US-10-732-923-3341	Sequence 3341, Ap
2	1549	62.3	316	3	US-09-837-344-31	Sequence 31, Appl
3	1549	62.3	316	3	US-09-900-963-31	Sequence 31, Appl
4	603	24.3	117	3	US-09-837-344-20	Sequence 20, Appl
5	603	24.3	117	3	US-09-900-963-20	Sequence 20, Appl
6	554	22.3	107	3	US-09-837-344-19	Sequence 19, Appl
7	554	22.3	107	3	US-09-900-963-19	Sequence 19, Appl
8	381.5	15.3	1003	5	US-10-732-923-3319	Sequence 3319, Ap
9	380.5	15.3	1407	5	US-10-732-923-3359	Sequence 3359, Ap
10	380.5	15.3	1549	5	US-10-732-923-3358	Sequence 3358, Ap
c 11	376.5	14.3	419	5	US-10-450-763-47660	Sequence 47660, A
12	374	15.0	1439	5	US-10-754-079-157	Sequence 157, App
13	360	14.5	350	4	US-10-101-487-58	Sequence 58, Appl
14	360	14.5	350	5	US-10-939-988-58	Sequence 58, Appl
15	359	14.4	1162	3	US-09-894-273-2	Sequence 2, Appli
16	359	14.4	1162	4	US-10-294-804-2	Sequence 2, Appli
17	359	14.4	1162	5	US-10-194-046-2	Sequence 2, Appli
18	337	13.6	1527	5	US-10-732-923-3354	Sequence 3354, Ap
19	331	13.3	2274	6	US-11-097-143-2763	Sequence 2763, Ap
20	323.5	13.0	611	3	US-09-216-393-81	Sequence 81, Appl
21	323.5	13.0	611	4	US-10-321-856-81	Sequence 81, Appl
22	320	12.9	658	6	US-11-097-143-23688	Sequence 23688, A
23	320	12.9	800	6	US-11-097-143-41169	Sequence 41169, A
24	318.5	12.8	1961	4	US-10-028-248A-103	Sequence 103, App
25	318.5	12.8	1961	4	US-10-107-782-103	Sequence 103, App
26	318	12.8	554	6	US-11-097-143-5154	Sequence 5154, Ap
27	317.5	12.8	1937	5	US-10-723-860-58	Sequence 58, Appl
28	317.5	12.8	1937	5	US-10-723-860-106	Sequence 106, App
29	315.5	12.7	1170	4	US-10-341-434-95	Sequence 95, Appl
30	315.5	12.7	1564	4	US-10-144-198-2	Sequence 2, Appli
31	315.5	12.7	1564	4	US-10-144-198-4	Sequence 4, Appli
32	315	12.7	647	5	US-10-691-672A-3	Sequence 3, Appli
33	315	12.7	1960	4	US-10-236-031B-62	Sequence 62, Appl
34	315	12.7	1960	4	US-10-028-248A-104	Sequence 104, App
35	315	12.7	1960	4	US-10-107-782-104	Sequence 104, App
36	315	12.7	1960	5	US-10-696-909A-49	Sequence 49, Appl
37	315	12.7	1960	6	US-11-009-554-27	Sequence 27, Appl
38	315	12.7	2143	5	US-10-450-763-32075	Sequence 32075, A
39	313.5	12.6	1959	4	US-10-028-248A-36	Sequence 36, Appl
40	313.5	12.6	1959	4	US-10-107-782-36	Sequence 36, Appl
41	312.5	12.6	1959	4	US-10-028-248A-106	Sequence 106, App
42	312.5	12.6	1959	4	US-10-107-782-106	Sequence 106, App
43	312.5	12.6	1985	5	US-10-732-923-3351	Sequence 3351, Ap
44	311.5	12.5	2056	6	US-11-097-143-4824	Sequence 4824, Ap
45	311	12.5	240	4	US-10-101-487-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-10-732-923-3341
; Sequence 3341, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3341
; LENGTH: 1909
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-3341

Alignment Scores:

Pred. No.:	4.53e-156	Length:	1909
Score:	2406.50	Matches:	487
Percent Similarity:	89.89%	Conservative:	2
Best Local Similarity:	89.52%	Mismatches:	4
Query Match:	96.80%	Indels:	51
DB:	5	Gaps:	2

US-09-837-344-41 (1-1482) x US-10-732-923-3341 (1-1909)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT---------------------------------- 87
              |||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405
```

```
Qy        88 ------------------GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                               ||||||||||||||||||||||||||||||||||||||||||
Db      1406 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 1425

Qy       130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
             ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1426 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 1445

Qy       190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1446 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 1465

Qy       250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 309
             ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db      1466 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 1485

Qy       310 GAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 369
             ||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1486 GluGlnGlnArgAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 1505

Qy       370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1506 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 1525

Qy       430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
             ||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||||
Db      1526 GluGlnGluArgLeuAlaAsnGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnGlu 1545

Qy       490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1546 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1565

Qy       550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------------ 597
             ||||||||||||||||||||||||||||||||||||||||||||||
Db      1566 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1585

Qy       597 ------------------------------------------------------------ 597

Db      1586 LeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlu 1605

Qy       598 ------------------------------GCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                                           ||||||||||||||||||||||||||||||
Db      1606 GlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArg 1625

Qy       628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1626 AspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGly 1645

Qy       688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
             |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1646 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 1665

Qy       748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1666 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 1685

Qy       808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1686 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 1705

Qy       868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1706 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 1725

Qy       928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1726 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 1745

Qy       988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1746 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 1765

Qy      1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1766 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 1785

Qy      1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1786 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 1805

Qy      1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1806 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 1825

Qy      1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1826 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 1845

Qy      1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1846 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 1865

Qy      1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1866 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 1885

Qy      1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1886 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 1905

Qy      1468 TTTATGAAACTA 1479
             ||||||||||||
Db      1906 PheMetLysLeu 1909
```

RESULT 2
US-09-837-344-31
; Sequence 31, Application US/09837344
; Patent No. US20020041882A1
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/837,344
; FILING DATE: 19-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/462,625
; FILING DATE: <Unknown>
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;          MOLECULE TYPE: peptide
;          PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;          SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-837-344-31

Alignment Scores:
Pred. No.:                1.36e-97        Length:          316
Score:                    1549.00         Matches:         314
Percent Similarity:       99.37%          Conservative:    0
Best Local Similarity:    99.37%          Mismatches:      2
Query Match:              62.31%          Indels:          0
DB:                       3               Gaps:            0

US-09-837-344-41 (1-1482) x US-09-837-344-31 (1-316)

Qy          13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy          73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy         133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy         193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy         253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy         313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy         373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy         433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
               ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy         493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy         553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy         613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy         673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy         733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy         793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy         853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy         913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 3
US-09-900-963-31
; Sequence 31, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 31:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-900-963-31

Alignment Scores:
Pred. No.:                1.36e-97        Length:          316
Score:                    1549.00         Matches:         314
Percent Similarity:       99.37%          Conservative:    0
Best Local Similarity:    99.37%          Mismatches:      2
Query Match:              62.31%          Indels:          0
DB:                       3               Gaps:            0

US-09-837-344-41 (1-1482) x US-09-900-963-31 (1-316)

Qy          13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
                 |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
                 |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                 ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 4
US-09-837-344-20
; Sequence 20, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 20:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 117 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-837-344-20

Alignment Scores:
Pred. No.:              4.35e-33        Length:         117
Score:                  603.00          Matches:        117
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            24.26%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-41 (1-1482) x US-09-837-344-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 5
US-09-900-963-20
; Sequence 20, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 20:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 117 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-900-963-20

Alignment Scores:
Pred. No.:              4.35e-33      Length:         117
Score:                  603.00        Matches:        117
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            24.26%        Indels:         0
DB:                     3             Gaps:           0

US-09-837-344-41 (1-1482) x US-09-900-963-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 6
US-09-837-344-19
; Sequence 19, Application US/09837344
; Patent No. US20020041882A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/837,344
;            FILING DATE: 19-Apr-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/462,625
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: FR 91 01286
;            FILING DATE: 05-FEB-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-078
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 19:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 107 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-837-344-19

Alignment Scores:
Pred. No.:              9.54e-30      Length:         107
Score:                  554.00        Matches:        107
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            22.28%        Indels:         0
DB:                     3             Gaps:           0
```

```
US-09-837-344-41 (1-1482) x US-09-837-344-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 7
US-09-900-963-19
; Sequence 19, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 19:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 107 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-900-963-19

Alignment Scores:
Pred. No.:              9.54e-30        Length:         107
Score:                  554.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            22.28%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-41 (1-1482) x US-09-900-963-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 8
US-10-732-923-3319
; Sequence 3319, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3319
;   LENGTH: 1003
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-732-923-3319

Alignment Scores:
Pred. No.:              5.35e-18        Length:         1003
Score:                  381.50          Matches:        101
Percent Similarity:     53.41%          Conservative:   126
Best Local Similarity:  23.76%          Mismatches:     105
Query Match:            15.35%          Indels:         93
DB:                     5               Gaps:           15

US-09-837-344-41 (1-1482) x US-10-732-923-3319 (1-1003)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              :::||||||          :::||||||   :::   :::|||   |||||||||   |||
Db        440 GlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProLeuGlnGluProGln 459

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                          |||   :::      :::::::::      :::||||||
```

```
Db        460 GlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGln 479

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              :::||||||         :::|||      |||:::  :::         :::::::::
Db        480 GlnGlnGluProGlnGlnGlnGluProGlnGlnGlnAspGluGlnGlnGlnAspGluGln 499

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
              ::::::   :::::::::      |||||||||  |||  |||||||||::::::
Db        500 GlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAsp---GluGlnGluGlnGlnAsp 518

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGAT------TTAGAACAAGATAGACTTGCTAAA 297
              :::::::::      |||||||||  |||        :::||||||        :::
Db        519 GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGluGln 538

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              ::::::   ||||||||||||   :::   ||||||:::         :::::::::|||
Db        539 GlnGluGluGlnGluGlnGlnGluGluGlnGluGlnGlnGluGluGlnGluGlnGluLeu 558

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              :::||||||:::   :::||||||::::::         :::::::::|||:::||||||
Db        559 GluGluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGln 575

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              :::   :::||||||::::::         :::::::::|||:::||||||:::   :::
Db        576 GluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGlu 592

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
              ||||||::::::         :::::::::|||:::||||||:::   :::||||||:::
Db        593 LeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu 609

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        610 Gln---------GluGlnGluLeuGluGluGluGlnGluLeuGluGluGln------ 624

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        625 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 643

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      |||::::::   |||      :::         |||:::
Db        644 GluGln---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGlu 662

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
              ||||||         ||||||      :::   :::
Db        663 LeuGluGluGlnGluGlnGluLeuGluGluGluGlnGluGln----------------- 676

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        676 ------------------------------------------------------------ 676

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              :::|||:::      :::|||
Db        677 LeuGluGluGlnGluGlnGlu---------------------------------------- 683

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                  |||||||||::::::         :::   ||||||:::::::::
Db        684 ---------LeuGluGluGlnGluGlnGluLeuGluGluGluGlnGluGlnGluGln--- 699

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                 :::   :::|||   :::
Db        700 ------------------------------------LeuGluGluValGluGluGln 706

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              :::      :::   |||   |||:::            :::                |||
Db        707 GluGlnGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGlu 726

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||   :::                        :::
Db        727 GluGlnGluGluGlnGluLeuGluGluValGlu----------------------Glu 738

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
                 ||||||              :::||||||||::::::   |||:::   |||
Db        739 GlnGluGlu--------------GlnGluLeuGluGluValGluGluGlnGluGln 753

Qy       1195 AATTTAGATGATTTAGATGAA-----------GGAATAGAAAAATCATCAGAAGAATTA 1242
                 |||:::::::::::::|||            |||:::|||:::      :::|||
Db        754 GluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGluThr 773

Qy       1243 TCTGAAGAAAAAATA 1257
                 ||||||   |||
Db        774 ValGluGluProIle 778

RESULT 9
US-10-732-923-3359
; Sequence 3359, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3359
;   LENGTH: 1407
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-10-732-923-3359

Alignment Scores:
Pred. No.:              6.23e-18        Length:         1407
Score:                  380.50          Matches:        127
Percent Similarity:     45.84%          Conservative:   132
Best Local Similarity:  22.48%          Mismatches:     167
Query Match:            15.31%          Indels:         139
DB:                     5               Gaps:           20

US-09-837-344-41 (1-1482) x US-10-732-923-3359 (1-1407)

Qy          1 CAAGAACAACAAAGCGATCTA---------GAACAAGAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||:::   :::|||         |||::::::|||:::   :::|||:::
Db        140 GlnGluGlnGluArgGluLeuAlaGluGluGluGluGlnArgLysLysArgGluArgPhe 159

Qy         52 CAAGAA-----------CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG 99
              ::::::           |||     |||   ||||||   ||||||   :::::::::
Db        160 GluGlnHisTyrSerArgGlnTyrArgAspLysGluGln---ArgLeuGlnArgGlnGlu 178

Qy        100 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 159
              |||:::|||:::::::::      |||:::|||:::|||   :::   |||   ::::::
Db        179 LeuGluGluArgArgAla------GluGluGluGlnLeuArgArgArgLysGlyArgAsp 196

Qy        160 CAACAAAGCGATCTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAA 210
                 :::      :::|||:::|||         ||||||   :::::::::|||:::
Db        197 AlaGluGluPheIleGluGluGluGlnLeuArgArgArgGluGlnGlnGluLeuLysArg 216

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG--------------TTG 255
              :::      :::   ||||||:::||||||   :::   :::                |||
Db        217 GluLeuArgGluGluGluGlnGlnArgArgGluArgArgGluGlnHisGluArgAlaLeu 236

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              ||||||:::::      |||   :::   |||       :::|||:::   :::||||||
Db        237 GlnGluGluGluGluGlnLeuLeuArgGlnArgArgTrpArgGluGluProArgGluGln 256

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
              |||         |||   ||||||       |||:::::::::|||::::::
Db        257 Gln-----------GlnLeuArgArgGluLeuGluGluIleArgGluArgGluGlnArg 272

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
              ||||||||||||   |||   :::::::::|||:::   :::|||      |||||||||
Db        273 LeuGluGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnArg---LeuGluGln 291

Qy        436 GAG--------------------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              |||                   |||         |||:::::::::|||::::::
Db        292 GluGluArgArgGluGlnGlnLeuArgArgGluLeuGluGluIleArgGluArgGluGln 311
```

```
Qy       475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
Db       312 ArgLeuGluGlnGluGluArgArgGluGlnArgLeuGluGlnGluGluArgArgGluGln 331

Qy       535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 594
Db       332 GlnLeuLysArgGluLeuGluGluIleArgGluArgGluGlnArgLeuGluGlnGluGlu 351

Qy       595 CGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
Db       352 ArgArgGluGlnLeuLeuAlaGluGluValArg-------------------------- 362

Qy       655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
Db       363 -----------GluGlnAlaArgGluArgGlyGluSerLeuThrArgArgTrpGlnArg 378

Qy       715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
Db       379 GlnLeuGlu---------SerGluAlaGlyAlaArgGlnSerLysValTyrSerArgPro 395

Qy       775 ---------CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825
Db       396 ArgArgGlnGluGluGlnSerLeuArgGlnAspGlnGluArgArgGlnArgGlnGluArg 415

Qy       826 GAAATATCTATAATAGAAAAAACAAATAGA---------------------------- 855
Db       416 GluArgGluLeuGluGluGlnAlaArgArgGlnGlnGlnTrpGlnAlaGluGluGluSer 435

Qy       856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT------ 909
Db       436 GluArgArgArgGlnArgLeuSerAlaArgProSerLeuArgGluArgGlnLeuArgAla 455

Qy       910 GAAGAAAAGAAAGATGGTTCAATAAAA------------CCAGAACAAAAAGAAGATAAA 957
Db       456 GluGluArgGlnGluGlnGluGlnArgPheArgGluGluGluGluGlnArgArgGluArg 475

Qy       958 TCTGCTGACATACAA---------------------------------------------- 972
Db       476 ArgGlnGluLeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGln 495

Qy       972 ---------------------------------------------------------- 972
Db       496 GlnLeuGlnGluGluAspSerPheGlnGluAspArgGluArgArgArgArgGlnGlnGlu 515

Qy       973 ----------------------------------------------------AATCAT 978
Db       516 GlnArgProGlyGlnThrTrpArgTrpGlnLeuGlnGluGluAlaGlnArgArgArgHis 535

Qy       979 ACATTA--------------GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
Db       536 ThrLeuTyrAlaLysProGlyGlnGlnGluGlnLeuArgGluGluGluGluLeuGlnArg 555

Qy      1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATAT---GACGATTCATTAATAGATGAAGAA 1080
Db       556 GluLysArgArgGlnGluArgGluArgGluTyrArgGluGluGluLysLeuGlnArgGlu 575

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db       576 GluAspGluLysArg-------------------------------------ArgArg 582

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db       583 GlnGluArgGluArgGlnTyrArgGluLeuGluGluLeuArgGlnGluGluGlnLeuArg 602

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
Db       603 AspArgLysLeuArgGluGluGluGlnLeuLeuGlnGluArgGluGluGluArgLeuArg 622

Qy      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT------TTTAAACCAAATGATAAA 1314
Db       623 ArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgGlnGluGluGln 642

Qy      1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
Db       643 GluLeuArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgArgGlu 662

Qy      1375 GAAAAATTCATAAAA 1389
Db       663 GluGlnGluLeuArg 667
```

RESULT 10
US-10-732-923-3358
; Sequence 3358, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3358
; LENGTH: 1549
; TYPE: PRT
; ORGANISM: Ovis aries
US-10-732-923-3358

Alignment Scores:

Pred. No.:	6.23e-18	Length:	1549
Score:	380.50	Matches:	133
Percent Similarity:	43.59%	Conservative:	129
Best Local Similarity:	22.13%	Mismatches:	194
Query Match:	15.31%	Indels:	145
DB:	5	Gaps:	18

US-09-837-344-41 (1-1482) x US-10-732-923-3358 (1-1549)

```
Qy         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db       197 GluArgGlnArgLeuGlnValGluGlnGlnGlnArgGlnArgGlyGluLeuArgGluArg 216

Qy        61 CAAAGCGATTTA--------------GAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
Db       217 GlnGluGluValGlnLeuGlnLysArgGluThrGlnGluLeuGlnArgGluArgLeuGlu 236

Qy       106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
Db       237 GluGluGlnGlnLeuGlnLysGlnLysArgGlyLeuGluGluArgLeuLeuGluGlnGlu 256

Qy       166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db       257 ArgArgGluGlnGluLeuArgArgLysGluGlnGluArgArgGluGlnGlnLeuArgGln 276

Qy       226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------------------- 264
Db       277 GluGlnGluGluAlaThrGlnGluGluIleSerGluArgGlyGluSerArgThrSerArg 296

Qy       265 ---CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG------------------ 303
Db       297 CysGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnArgLysValTyrSerArgPro 316

Qy       304 ---TTACAAGAGCAGCAAAGC----------------------GATTTAGAACAAGAG 336
Db       317 HisArgGlnGluGlnGlnSerArgArgGlnGluGlnGluLeuLeuGluArgGlnGlnGlu 336

Qy       337 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------------TTAGAA 381
Db       337 GlnGlnIleSerGluGluValGlnSerLeuGlnGluAspGlnGlyArgGlnArgLeuLys 356

Qy       382 CAAGAGAGACGT---------------------------------------------- 393
Db       357 GlnGluGlnArgTyrAspGlnAsnTrpArgTrpGlnLeuGluGluGluSerGlnArgArg 376
```

```
Qy        394 --------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
                            ||||||         :::||||||      :::   ||||||
Db        377 ArgTyrThrLeuTyrAlaLysProAlaGlnArgGluGlnValArgGluGluGluGlnLeu 396

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA------ 492
              ||||||   :::|||||||||||||   ::::::       |||:::||||||
Db        397 ArgLeuLysGluGluLysLeuGlnArgGluLysArgArgGlnGluArgGluArgGlnTyr 416

Qy        493 ------------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                               |||   :::|||:::||||||   ::::::      |||:::
Db        417 ArgGluValGluLeuGlnArgGluGluGluArgLeuGlnArgGluGluGluGlnLeuGln 436

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA-------------------------- 567
              :::|||   |||   |||   :::   ||||||
Db        437 ArgGluGluArgGluLysArgArgArgGlnGluArgGluLysGlnTyrLeuGluLysVal 456

Qy        568 ---------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
                      ::::::      |||:::::|||   |||   |||   :::   :::
Db        457 GluLeuTrpGluGluGluGlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArg 476

Qy        619 CAGCAAAGAGATTTAGAA----------------------------CAAAGGAAGGCT 648
              :::::::::   ||||||                            ||||||:::
Db        477 GluLysGlnTyrLeuGluLysValGluLeuArgGluGluGluGlnLeuGlnArgGlnGlu 496

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAG---------GAACATGGAGATATATTAGCA 699
                  |||:::      ||||||::::::         |||      ::::::
Db        497 ArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGlu 516

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||:::      |||   |||         :::         |||   ||||||
Db        517 GluGluGlnLeuGlnArgGluGlu------ArgGluLysArgArgGlnGluArgGluArg 534

Qy        760 GGATATTAT------------ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
                |||               :::      :::      |||   :::      |||
Db        535 GlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGlu 554

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA--------- 858
                  |||      :::|||         :::||||||         :::
Db        555 LysArgArgGlnGluArgGluLysGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 574

Qy        859 ------TCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                        :::   |||::::::      |||   :::||||||
Db        575 GlnLeuGlnArgGlnGluArgGlnLysArgArgGlnGluArgGluLysGlnTyrLeuGlu 594

Qy        913 GAA---------AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
              :::         :::::::::      ::::::   :::::::::|||   :::
Db        595 LysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGluLysArgArgGln 614

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA--- 1020
              :::   :::         ||||||   |||   :::   :::                |||
Db        615 GluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnValGlnArg 634

Qy       1021 ------------------------------------------------------ATAAGT 1026
                                                                    :::
Db        635 GlnGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysGluLeuGln 654

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              :::   |||:::   :::      |||      :::   |||:::   ||||||   :::
Db        655 ArgGlnGluGluArgLeuGlnGluGlu------GluGlnLeuLeuArgGluGluArgGlu 672

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                        :::      :::      ::::::         |||:::||||||
Db        673 LysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 692

Qy       1147 AACATA-------------------------------------GGAATTTATAAAGAA 1167
                 :::                                     |||   |||
Db        693 GlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGlu 712

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
                 |||:::|||      :::   |||   |||             |||   :::::::::
Db        713 LysGluGluLeuGlnArgGlnGluGluArgLeuGlnArgGluLysGluGlnLeuGlnArg 732

Qy       1228 TCATCAGAAGAA----------------------------TTATCTGAAGAAAAAATA 1257
                                                    |||      :::|||||||||
Db        733 GluAspArgGluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGluLeuGln 752

Qy       1258 AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
                    ::::::      :::   ::::::   |||          :::      :::|||
Db        753 GlnGluGluAspArgLeuGlnArgGluLysGlnLeuLeuArgGluAspArgGluLysArg 772

Qy       1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
                  |||   |||   ::::::      :::   ::::::|||:::   :::||||||
Db        773 GlnTyrLeuGluLysValGluLeuGlnArgGluGluGluGlnLeuGlnArgGluLysArg 792

Qy       1378 AAA 1380
                 :::
Db        793 Arg 793
```

RESULT 11
US-10-450-763-47660
; Sequence 47660, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 47660
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (362)..(408)
; OTHER INFORMATION: Vertebrate metallothioneins proteins domain identified by
; OTHER INFORMATION: eMATRIX, accession number BL00203, p-value=2.484e-11, raw score
; OTHER INFORMATION: 13.94
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(419)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-47660

Alignment Scores:

Pred. No.:	1.19e-17	Length:	419
Score:	376.50	Matches:	149
Percent Similarity:	44.34%	Conservative:	35
Best Local Similarity:	35.90%	Mismatches:	76
Query Match:	14.26%	Indels:	155
DB:	5	Gaps:	25

US-09-837-344-41 (1-1482) x US-10-450-763-47660 (1-419)

```
Qy       1217 CCTTCATCTAAATCATCTAAATTTTCATTTTTCTCTATCAAATCTTCTAGTTCTTTATAA 1158
              |||:::      ||||||      |||      |||:::   ||||||   ||||||
Db        136 ProAlaProAlaSerSerProCysSer---ArgSerLeuSerSerSerArgSerLeu--- 153

Qy       1157 ATTCCTATGTTTTCTTCATCTTGGAAA-----------------TTGTCATATTGCACA 1116
                        |||||||||   :::                  ||||||   |||
Db        154 ------------SerSerSerCysArgValSerCysCysHisSerLeuSerSerCysCys 169

Qy       1115 ATAGGCTTAAATTCGTCTAAGTCTTCATCATCTTCTTCTTCATCTATTAATGAATCGTCA 1056
                    |||:::   |||:::|||||||||:::||||||   |||   :::   ||||||
Db        170 ArgSerLeuSerCysSerArgSerSerSerThrSerSerArgSerSerSerThrSerSer 189

Qy       1055 TATTCAGCACTTATTTCATCCTCATACTTACTTATTTGAAAATCATTAACATCAGAAATA 996
                 |||      :::|||||||||
```

```
Db        190 SerSerLeuSerLeuSerSerSerLeu-------------------------------- 198

Qy        995 TTTACTGTCTCTAATGTATGATTTTGTATGTCAGCAGATTTATCTTCTTTTTGTTCTGGT 936
                                  |||   |||      ||||||      ||||||
Db        199 ---------------------SerCysSerSerCysArgLeuSer------CysSer--- 208

Qy        935 TTTATTGAACCATCTTTCTTTTCTTCAAGATGTCCTTTATGTATATCCCTTCGTCCTTCA 876

Db        208 ------------------------------------------------------------ 208

Qy        875 ACATTTGTTGTAATAGATTCTCTATTTGTTTTTTCTATTATAGATATTTCCTTGGAATCT 816

Db        208 ------------------------------------------------------------ 208

Qy        815 CTACTATTCCCTCTGTTGTCCTGAGGTAAAGAAGATTGATGTGGTATATAATATCCACGT 756

Db        209 ------------------------------------------------------------Asn 209

Qy        755 TCATTTTCTGATGGAAGTTCTATAGCTGGTATTTCTAAACGACCATATAAATCCTCTGCT 696
              |||||||||      ||||||         :::|||
Db        210 SerPheSerThrSerSerSerSerLeuSerLeuSer---------------------- 221

Qy        695 AATATATCTCCATGTTCCTTTTTTCTTTCTAAATTTTTTTTCGTATCAGCCTTCCTTTGT 636
                                                  |||:::      |||
Db        222 ---------------------------------------------SerSerLeuSerCys 226

Qy        635 TCTAAATCTCTTTGCTGCTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCG 576
              |||   |||   |||   ||||||:::   |||   :::      |||   |||   |||
Db        227 Ser---SerCysCysLeuSerCysSerAsnSerPheSerThrSerSerSerSerLeuSer 245

Qy        575 CTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGT 516
              |||       |||   :::   |||       ||||||||||||||   |||
Db        246 LeuSerSerSerLeuSerCysSerSerCysArgLeuSerCysSerAsnSerPhe------ 263

Qy        515 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAAC 456
                  :::   |||   :::   |||||||||      |||      ||||||   :::
Db        264 ------SerThrSerSerSerSerLeuSerCysProGlySerPheSerCysSerHisSer 281

Qy        455 TTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTA 396
                 |||           |||||||||:::||||||      ||||||:::   ||||||
Db        282 LeuSer-----------SerCysSerArgSerLeu---SerSerCysSerHisSerLeu 296

Qy        395 GCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTC 336
              :::      |||||||||:::||||||      ||||||:::   ||||||:::
Db        297 Ser------SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------ 311

Qy        335 TCTTGTTCTAAATCGCTTTGCTGCTCTTGTAACTTTTCTTTAGCAAGTCTATCTTGTTCT 276
              |||||||||:::||||||      ||||||:::   ||||||:::      |||||||||
Db        312 SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSer 328

Qy        275 AAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTT 216
              :::||||||      ||||||:::   ||||||:::      |||||||||:::||||||
Db        329 ArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSerArgSerLeu 345

Qy        215 ---TGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTTTGTTGT 159
                 |||||||||:::   |||            ||||||||||||||||||
Db        346 SerSerCysSerCysSerLeuSer------------SerCysSerArgSerLeu---Ser 360

Qy        158 TCTTGCAACTTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGCTGCTCTTGTAAC 99
              ||||||:::   ||||||:::      |||||||||   ||||||      ||||||:::
Db        361 SerCysSerHisSerLeuSer------SerCysSerHisSerLeu---SerSerCysSer 377

Qy         98 TTTTCTTTAGCAAGTCTATCTTGTTCTAAATCGCTTTGT------------------TGT 57
                 ||||||:::|||   |||||||||   |||   |||                   |||
Db        378 ArgSerLeuSerSerCysSerCysSerLeuSerSerCysSerArgCysLeuSerSerCys 397

Qy         56 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTT 12
              ||||||:::   |||            ||||||||||||||||||
Db        398 SerCysSerLeuSer------------SerCysSerArgSerLeu 408

RESULT 12
US-10-754-079-157
; Sequence 157, Application US/10754079
```

```
; Publication No. US20050008602A1
; GENERAL INFORMATION:
;  APPLICANT: Hirai, Yohei
;  APPLICANT:  Takebe, Kyoko
;  TITLE OF INVENTION: OLIGOPEPTIDES FOR PROMOTING HAIR GROWTH
;  FILE REFERENCE: 467392000222
;  CURRENT APPLICATION NUMBER: US/10/754,079
;  CURRENT FILING DATE:  2004-01-07
;  PRIOR APPLICATION NUMBER: US 10/155,922
;  PRIOR FILING DATE: 2002-05-23
;  PRIOR APPLICATION NUMBER: PCT/JP01/04691
;  PRIOR FILING DATE: 2001-06-04
;  PRIOR APPLICATION NUMBER: JP 2000-166903
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: JP 2001-347340
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: JP 2001-347338
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: JP 2001-371175
;  PRIOR FILING DATE: 2001-12-05
;  PRIOR APPLICATION NUMBER: JP 2001-371336
;  PRIOR FILING DATE: 2001-12-05
;  PRIOR APPLICATION NUMBER: JP 2003-1891
;  PRIOR FILING DATE: 2003-01-08
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157
;   LENGTH: 1439
;   TYPE: PRT
;   ORGANISM: Mus Musculus
US-10-754-079-157

Alignment Scores:
Pred. No.:              1.73e-17        Length:         1439
Score:                  374.00          Matches:        143
Percent Similarity:     48.15%          Conservative:   130
Best Local Similarity:  25.22%          Mismatches:     170
Query Match:            15.04%          Indels:         124
DB:                     5               Gaps:           25

US-09-837-344-41 (1-1482) x US-10-754-079-157 (1-1439)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------------ 48
                ::|||::::::   :::|||||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||      |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                    ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                    ||||||::||||||      :::   :::   :::|||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                ||||||||||||      ::|||:::         |||||||||::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||          |||   :::   ||||||::::::         |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              :::::::|||   :::::::::   :::::::|||:::         ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755
```

```
Qy      391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
Db      756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy      442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
Db      776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy      493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
Db      796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy      544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db      816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy      598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db      836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy      658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
Db      854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy      718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db      871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy      778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT------------------ 819
Db      887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy      819 ------------------------------------------------------------ 819

Db      907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy      820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
Db      927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy      865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db      947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy      925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
Db      965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy      979 ----------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
Db      982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy     1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
Db     1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy     1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
Db     1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGlnLeuArg 1040

Qy     1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
Db     1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy     1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
Db     1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy     1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
Db     1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy     1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
Db     1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy     1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
Db     1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy     1357 CAGGTTAATAAGGAAAAGGAA 1377
Db     1141 LysPheArgGluGluGlnGlu 1147

RESULT 13
US-10-101-487-58
; Sequence 58, Application US/10101487
; Publication No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 350
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-58

Alignment Scores:
Pred. No.:              1.59e-16        Length:         350
Score:                  360.00          Matches:        78
Percent Similarity:     50.24%          Conservative:   132
Best Local Similarity:  18.66%          Mismatches:     140
Query Match:            14.48%          Indels:         68
DB:                     4               Gaps:           2

US-09-837-344-41 (1-1482) x US-10-101-487-58 (1-350)

Qy        1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        1 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20

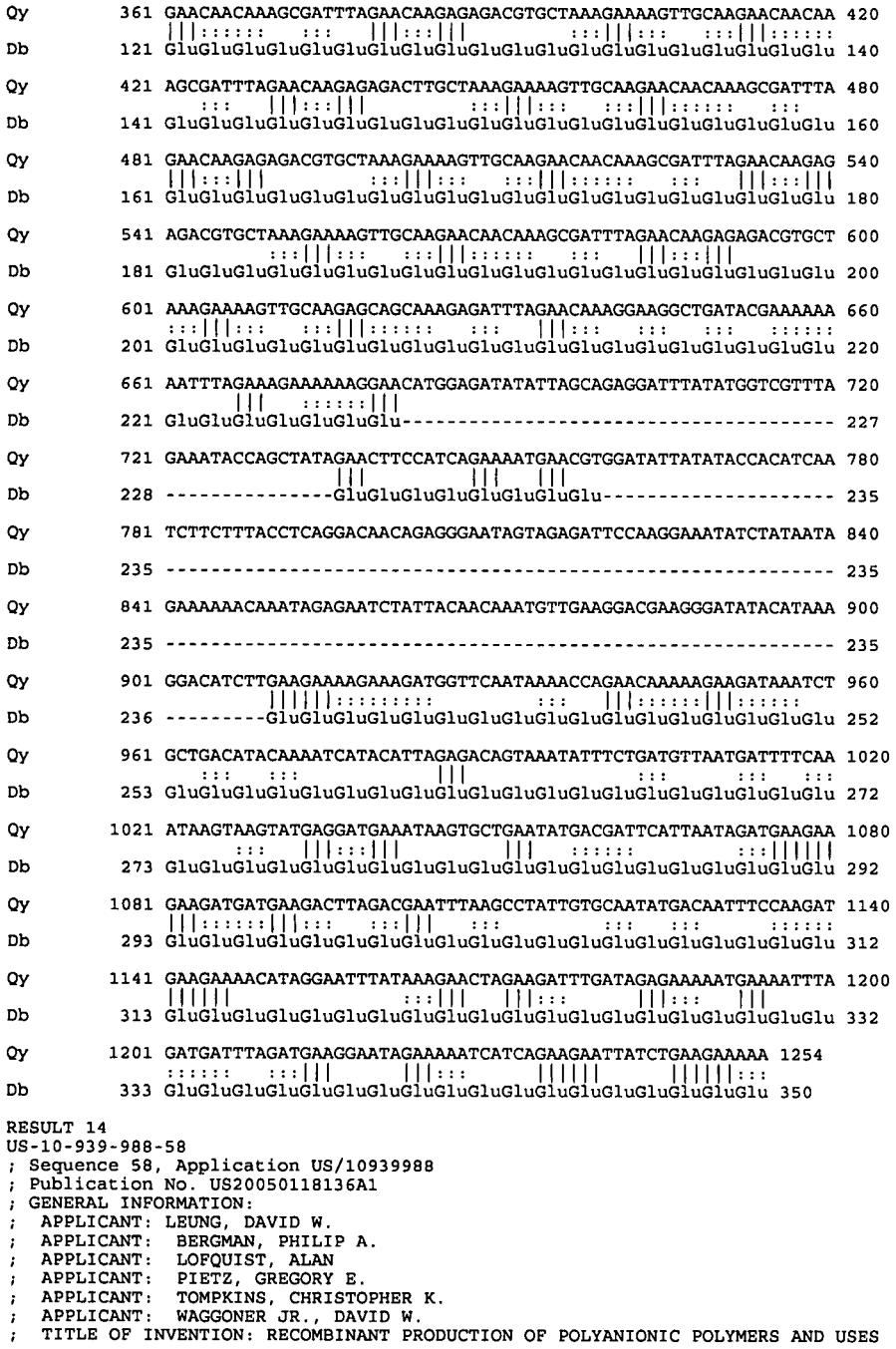
Qy       61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db       21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40

Qy      121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db       41 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 60

Qy      181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db       61 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80

Qy      241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db       81 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100

Qy      301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db      101 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
```

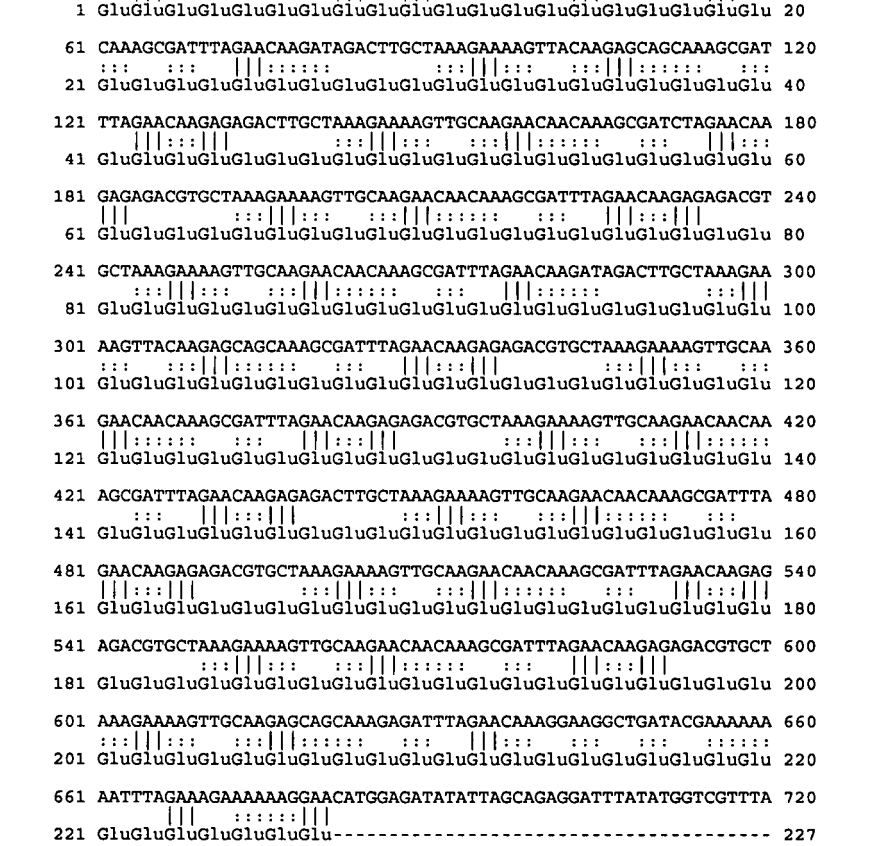


RESULT 14
US-10-939-988-58
; Sequence 58, Application US/10939988
; Publication No. US20050118136A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/939,988
; CURRENT FILING DATE: 2004-09-14
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-939-988-58

Alignment Scores:

Pred. No.:	1.59e-16	Length:	350
Score:	360.00	Matches:	78
Percent Similarity:	50.24%	Conservative:	132
Best Local Similarity:	18.66%	Mismatches:	140
Query Match:	14.48%	Indels:	68
DB:	5	Gaps:	2

US-09-837-344-41 (1-1482) x US-10-939-988-58 (1-350)



```
Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db        228 --------------GluGluGluGluGluGluGlu---------------------- 235

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
Db        235 ------------------------------------------------------------ 235

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db        235 ------------------------------------------------------------ 235

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db        236 ---------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        253 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 272

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        273 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 292

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db        293 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 312

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db        313 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 332

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
Db        333 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 350
```

RESULT 15
US-09-894-273-2
; Sequence 2, Application US/09894273
; Publication No. US20040037847A1
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/894,273
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/109,422
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

Alignment Scores:

Pred. No.:	1.83e-16	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	49.44%	Conservative:	127
Best Local Similarity:	21.03%	Mismatches:	110
Query Match:	14.44%	Indels:	116
DB:	3	Gaps:	12

US-09-837-344-41 (1-1482) x US-09-894-273-2 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
Db        809 GlnGluLeuGluGluGlnGluGlnGlu------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db        817 ------------------------------------------------------------ 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
Db        817 ------------------------------------------------------------ 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
Db        838 ---------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876
```

```
Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
              :::|||:::                    |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                |||:::::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                ::::::   ::::::   |||:::   :::            :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                    :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 16
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE: UM-03778
;  CURRENT APPLICATION NUMBER: US/10/294,804
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: US/09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

Alignment Scores:
Pred. No.:                1.83e-16        Length:         1162
Score:                    359.00          Matches:        94
Percent Similarity:       49.44%          Conservative:   127
Best Local Similarity:    21.03%          Mismatches:     110
Query Match:              14.44%          Indels:         116
DB:                       4               Gaps:           12

US-09-837-344-41 (1-1482) x US-10-294-804-2 (1-1162)

Qy           4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
               |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db         604 GluGlnGlnGlnAsp---GluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGln 622

Qy          64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                  :::   :::||||||          :::::::::   |||      |||||||||
Db         623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGln 642

Qy         118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
               |||   :::|||:::          ::::::      |||:::|||            :::
Db         643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGln 662

Qy         178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                  |||::::::      |||:::   |||::::::|||          ||||||::::::
Db         663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGln 682

Qy         238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||          :::
Db         683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGluGln 702

Qy         298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
               :::      ||||||||||||   :::   ||||||:::         ::::::
Db         703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy         358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
                  |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db         723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy         418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
               |||   |||   :::|||:::            :::      ||||||||||||   :::
Db         743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy         478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                  ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db         759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy         538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
               :::         :::::::::|||:::||||||:::   :::||||||::::::
Db         779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy         598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                  :::::::::|||:::||||||:::::::::||||||::::::             :::
Db         794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------------Glu 808

Qy         658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
               :::   ||||||   :::::::::
Db         809 GlnGluLeuGluGluGlnGluGlnGlu-------------------------------- 817

Qy         718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db         817 ------------------------------------------------------------ 817

Qy         778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db         817 ------------------------------------------------------------ 817

Qy         838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
               :::|||:::      :::|||
Db         818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy         898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                        |||||||||::::::            :::   ||||||
Db         825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy         958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                              ::::::   :::
Db         838 ---------------------------------------------GluLeuGluGluGln 842

Qy        1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
               ::::::   :::   |||   |||:::      :::                  ::::::
Db         843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy        1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
               |||:::::::::::::::|||:::|||   :::         |||              |||
Db         857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy        1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
               :::|||:::                    |||||||||::::::   |||:::   |||
Db         877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy        1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                  |||:::::::::::::|||      |||:::      ||||||:::   |||::::::
Db         891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy        1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                  ::::::   ::::::   |||:::   :::            :::      ::::::
Db         911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy        1315 AGTTTGTATGATGAGCATATT 1335
                      :::|||   |||
Db         931 GluThrValGluGluProIle 937

RESULT 17
US-10-194-046-2
; Sequence 2, Application US/10194046
; Publication No. US20040248081A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
```

```
;  TITLE OF INVENTION: LANA Viral Protein Binding Sites
;  FILE REFERENCE: UM-07142
;  CURRENT APPLICATION NUMBER: US/10/194,046
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-194-046-2

Alignment Scores:
Pred. No.:                1.83e-16      Length:          1162
Score:                    359.00        Matches:           94
Percent Similarity:       49.44%        Conservative:     127
Best Local Similarity:    21.03%        Mismatches:       110
Query Match:              14.44%        Indels:           116
DB:                       5             Gaps:              12

US-09-837-344-41 (1-1482) x US-10-194-046-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::     :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                :::   :::||||||        :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                 |||::::::      |||:::   |||:::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||        :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||::::::            :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln-----------Glu 808

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   ||||||   :::::::::
Db        809 GlnGluLeuGluGluGlnGluGlnGlu---------------------------------- 817

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
```

```
Db        817 ---------------------------------------------------------- 817

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        817 ---------------------------------------------------------- 817

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                  :::|||:::      :::|||
Db        818 LeuGluGluGlnGluGlnGlu---------------------------------------- 824

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                        |||||||||::::::            :::   ||||||
Db        825 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGln------------ 837

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                                                          ::::::   :::
Db        838 ----------------------------------------------GluLeuGluGluGln 842

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                 ::::::   :::   |||   |||:::      :::                  ::::::
Db        843 GluValGluGluGlnGluGlnGluValGluGluGln-----------------GluGln 856

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
                 |||:::::::::::::::|||:::|||   :::          |||             |||
Db        857 GluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGln 876

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
                 :::|||:::                |||||||||::::::   |||:::   |||
Db        877 GluGluGln-----------------GluLeuGluGluValGluGluGlnGluGluGln 890

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                  |||:::::::::::|||      |||:::      ||||||:::   |||::::::
Db        891 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGlu 910

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                 ::::::   ::::::   |||:::   :::            :::      ::::::
Db        911 GlnGlnGluLeuGluGluValGluGluGlnGluGlnGlnGlyValGluGlnGlnGluGln 930

Qy       1315 AGTTTGTATGATGAGCATATT 1335
                      :::|||   |||
Db        931 GluThrValGluGluProIle 937

RESULT 18
US-10-732-923-3354
; Sequence 3354, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3354
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Plasmodium yoelii yoelii
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1527)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3354

Alignment Scores:
Pred. No.:                5.76e-15      Length:          1527
Score:                    337.00        Matches:          135
Percent Similarity:       45.21%        Conservative:     134
Best Local Similarity:    22.69%        Mismatches:       190
Query Match:              13.56%        Indels:           136
DB:                       5             Gaps:              23
```

US-09-837-344-41 (1-1482) x US-10-732-923-3354 (1-1527)

```
Qy          1 CAAGAACAA------CAAAGCGATCTAGAACAA---GAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||      :::   ::::::::::   |||   :::   |||||||||:::
Db        755 GlnGluGlnMetLeuLysAspGluIleGlnLysIleGluLeuLysThrLysGluLysIle 774

Qy         52 CAAGAACAACAAAGCGATTTAGAACAA-----------------GATAGACTTGCTAAA 93
              :::|||   :::   |||   |||:::                   :::   |||
Db        775 GluGluLeuLysGlnAspPheGluLysThrGlnLysIleAsnMetGluAsnLeuGluMet 794

Qy         94 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG---AGACTTGCTAAAGAAAAG 150
              ||||||                :::::|||||   |||          |||||||||
Db        795 GluLysGluSerPheIleAsnAsnAsnLeuGluAsnGluIleAsnLysMetLysGluLys 814

Qy        151 TTGCAAGAA-----------------CAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||:::|||                ::::::::::::::::::   :::   :::
Db        815 LeuGluGluLysTyrGluThrGlnIleLysGluThrGluMetLysTyrLysTyrGlnIle 834

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA---GAA 249
              ||||||:::::::::::::   :::   :::   ||||||          :::|||   |||
Db        835 LysGluGluIleGluLysThrLysGlnAsnAlaGluGlnAsnPheAsnSerLysPheGlu 854

Qy        250 AAGTTGCAAGAACAA------------------------------------------- 264
              |||   :::|||
Db        855 LysTyrLysGluAsnLeuGluLysAsnLysAsnAspPheIleAsnAsnLeuIleIleGlu 874

Qy        265 ----------------------CAAAGCGATTTAGAACAAGATAGACTTGCTAAA--- 297
                                   ::::::|||:::||||||   :::      :::
Db        875 LysAsnAsnGluIleGluSerPheLysAsnAspIleGluGlnLysLysPheLysGluMet 894

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA---------GAGAGACGTGCTAAA 348
              ||||||   :::|||:::          |||   |||          |||
Db        895 GluLysPheLysGluGluAsnGluLysLeuLeuGlnAsnAsnPheGluAsnMetLysAsn 914

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
                  ::::::||||||       :::|||     ::::::     |||   :::
Db        915 HisPheIleGluGluGlnAsnLysHisIleGluAsnIleLysLysGluTyrGluLeuIle 934

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
              :::      :::   :::       :::|||   :::   |||   |||:::||||||
Db        935 LysAsnAsnGluIleGluTyrLeuLysGluGluMetLysLysAsnLysIleGlnGluIle 954

Qy        469 CAAAGC-----------------------------------------GATTTAGAACAA 486
              ::::::                                        |||:::::::::
Db        955 GluAsnValGluLeuLysLeuAlaAspGluLysAsnLysHisIleAspAspMetLysLys 974

Qy        487 GAG--------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA 519
              |||                         |||||||||:::      :::
Db        975 GluLeuGluAsnIleTyrAsnValGluIleAsnAsnLeuLysGluLysIleAsnAlaGlu 994

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||:::                   |||:::|||   |||:::   |||:::::::::
Db        995 GlnThr-----------------LeuAlaGlnGluMetLeuGluThrGlnArgAsnGlu 1008

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                 ||||||:::        :::            |||||||||   ::::::   :::
Db       1009 TyrGluGlnLysLeuGluIleGlnArgAsnGluTyrGluGlnGlnLeuGluIleArgArg 1028

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              :::   :::   |||      :::|||::::::||||||      |||:::   :::
Db       1029 AsnGluMetGluAsnLysLeu---IleGluGlnArgLysGluSerTyrAspLeuGlySer 1047

Qy        700 GAGGAT-----------------TTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
              :::                   |||      ::::::|||          :::|||:::
Db       1048 LysLysGluSerGluIleGlnIleLeuThrAspLysIleGluLysLeuLysMetGluIle 1067

Qy        742 CCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAAC 801
                  |||:::            :::   |||      |||:::     |||
Db       1068 GluIleIleAsnLysGluLysIleAspLeuGluHisAsnIleSerIleLeuAsnAspGlu 1087

Qy        802 AGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCT 861
              |||   |||      :::                                  ::::::
```

```
Db       1088 ArgGluAsnIleLeuAsn----------------------------GlnAsnLysAsn 1097

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACAT----------------------- 897
              :::      |||::::::        ||||||:::
Db       1098 LeuGluAspAsnLeuLysAsnAsnGluAspIleTyrAsnLysLysIleAsnIleLeuAsn 1117

Qy        898 ------AAAGGACATCTTGAAGAA---------------AAGAAAGATGGTTCAATAAAA 936
                    |||      ||||||:::               |||::::::|||:::   :::
Db       1118 AspGluLysLeuLysLeuGluLysGluIleGluHisIleLysGluAsnGlyAlaLysGlu 1137

Qy        937 CCAGAACAAAAAGAAGATAAATCTGCTGAC-----------------------ATACAA 972
                 :::|||      ::::::   ||||||                        |||:::
Db       1138 SerAspGlnIleArgGluGlnPheAlaAspLeuLeuGlnGluGluIleAspArgIleLys 1157

Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
                    :::   |||   |||||||||                    :::|||::::::
Db       1158 LysGluSerLysGluLysValAsnIleTyrIleGlnGlnTyrAsnGluIleAsnGluGlu 1177

Qy        1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAA 1092
               :::|||          ||||||:::|||   |||                  ::::::
Db       1178 TyrGluGluLysLysLysGluTyrAsnAspLeuLeuGluLysAlaAsnValSerAsnLys 1197

Qy        1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
                  |||   |||                        :::   ||||||||||||
Db       1198 GlnLeuThrGlu----------------------------LysCysGluGluAsnIle 1207

Qy        1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT 1212
                       |||   ||||||:::|||:::      ||||||      ::::::
Db       1208 GlnLysIleAsnGluTyrGluAspMetIleLysMetLeuGluAsnGlnThrGluVal--- 1226

Qy        1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAA 1272
               :::   |||      :::||||||:::|||:::   :::||||||   :::
Db       1227 ---LeuValThrLysLysIleGlnGluLeuAsnGluAspPheLeuLysLysLysGluAla 1245

Qy        1273 TATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT---GAG 1329
               ::::::      |||::::::      |||      :::   ::::::   :::   :::
Db       1246 PheAspAsnGluLysAsnAspLeuLeuLysAsnTyrGluHisAlaIleThrGluAsnLys 1265

Qy        1330 CATATTAAA---AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
               |||||||||   :::   :::|||           |||:::||||||      |||
Db       1266 HisIleLysGluGlnLeuGluAsnPheThrAsnSerAsnGluGluLysIleSerGlnIle 1285

Qy        1387 AAATCATTGTTT--------------------------CATATATTTGACGGAGACAAT 1419
               |||:::   :::                         |||      :::      |||
Db       1286 LysAsnGlnTyrGluThrGlnIleLysGluMetGlnIleHisPheAspGluLeuLeuAsn 1305

Qy        1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
               |||                :::|||:::   |||||||||   |||
Db       1306 GluThrLysAsnLysAsnAsnGluMetLeuGluAspIleValLys 1320
```

RESULT 19
US-11-097-143-2763
; Sequence 2763, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28

```
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2763
;   LENGTH: 2274
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-2763
```

Alignment Scores:

Pred. No.:	1.47e-14	Length:	2274
Score:	331.00	Matches:	134
Percent Similarity:	44.30%	Conservative:	134
Best Local Similarity:	22.15%	Mismatches:	167
Query Match:	13.31%	Indels:	170
DB:	6	Gaps:	27

US-09-837-344-41 (1-1482) x US-11-097-143-2763 (1-2274)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTG 51
              :::|||::::::      |||      ||||||      |||          ||||||:::
Db       1385 LysGluArgGluGluSerLeuArgMetGluArgGluGluLysLeuLysGluGluLysIle 1404

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
              :::|||::::::           :::::::::|||   :::||||||:::::::|||:::
Db       1405 LysGluLysGlu---------ArgGluGluLysLeuLysGluGluLysIleLysGluLys 1421

Qy        112 CAAAGCGAT------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
              |||    :::      |||:::::::|||:::|||   :::::::::   :::|||:::
Db       1422 GlnArgGluGluLysLeuLysGluGluLysLeuLysGluLysGluArgGluGluArgMet 1441

Qy        166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA--------- 216
                 :::   |||:::|||   :::|||||||||||   :::|||:::::
Db       1442 LysGluLysGluArgGluGluLysAlaLysGluLysGlnArgGluGluLysLeuArgGlu 1461

Qy        217 ---------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                      :::   |||:::|||   :::   ||||||||||||:::|||::::::
Db       1462 GluLysIleLysGluLysGluArgGluGluLysLeuLysGluLysLeuArgGluGluLys 1481

Qy        268 ---AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
                 :::   |||::::::          :::::::::   :::|||:::      :::
Db       1482 IleLysGluLysGluLysGluGluLysLeuArgLysGluArgGluGluLysMetArgGlu 1501

Qy        325 TTAGAACAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAACAA 369
                 |||:::|||   :::   |||                ||||||::::::|||::::::
Db       1502 LysGluArgGluGluLysIleLysGluLysGluArgValGluLysIleLysGluLysGlu 1521

Qy        370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                ::::   :::   ::::::      :::||||||||:::|||::::::      |||
Db       1522 ArgGluGluLysLeuLysLysGluLysGluGluLysLeuLysGluLysGluGluLeuLeu 1541

Qy        430 GAA--------------------------CAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::                          :::|||:::|||   :::::::::   |||
Db       1542 LysLysLysGluLysGluGluLysGluArgGluGluLysLeuLysGluLysGluArgGln 1561

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||:::      :::   |||:::|||   :::   |||   :::   :::|||:::|||
Db       1562 GluLysLeuLysGluLysGluArgGluGluLysLeuLysArgGluThrGluGluArgGln 1581

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGAT 579
                 :::   |||:::|||   :::   |||||||||   :::   |||:::      |||
Db       1582 ArgGluLysGluArgGluGluLysLeuLysGluLysGluArgAlaGluLysLeuLysAsp 1601

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAGCAGCAAAGAGAT 630
              ||||||:::|||   :::   |||||||||          |||:::|||:::::::::::
Db       1602 LeuGluLysGluValLysLeuLysGluLysGluGluGlnLeuLysGluLysGluLysGlu 1621

Qy        631 TTA---------GAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
                          |||      :::|||
Db       1622 LeuLysLeuLysGluLysLysGluLysAspLysValLysGluLysGluLysSerLeuGlu 1641

Qy        682 CATGGAGATATATTA----------------------------------GCAGAGGAT 705
                    :::|||                                    ||||||
Db       1642 SerGluLysLeuLeuIleSerAlaThrValSerAsnProTrpArgArgValValGluAsp 1661

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATA---GAACTTCCATCAGAAAATGAACGTGGA 762
                    :::|||      ||||||:::   :::   ||||||      ::::::
Db       1662 ThrProProLysLeu------ProAlaValGlnAspTyrProSerLeuGlyLysLys--- 1678

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA----------------- 804
                    |||   :::::::|||   |||:::      |||
Db       1679 ---------ProThrLysAlaSer---ProGluLysLysArgAspGluLysLeuLeuPro 1694

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATATCT---------------------------- 834
              |||   :::          ||||||::::::
Db       1695 GlyLeuThrThrProProLysGluValAsnAspThrPheGluAspPheLeuSerGlyLeu 1714

Qy        835 ------------------------------------------------ATAATAGAAAAA 846
                                                           :::   |||
Db       1715 LysProLeuGluAlaLeuProProLeuProAlaLeuGluProLeuGluValLysGluAsp 1734

Qy        847 ACAAATAGAGAATCTATTACA---ACAAATGTTGAAGGA---------------------- 882
              :::   :::||||||::::::      |||   |||
Db       1735 SerLysLysGluSerValSerLeuIleAsnPheGluSerProLeuGlnGluThrAlaSer 1754

Qy        883 ---CGAAGGGATATACAT---------AAAGGACATCTTGAAGAAAAG------------ 918
                 |||   |||              :::|||     |||:::
Db       1755 IleProArgLysIleSerProProProArgGlyPheThrGluGlnAsnLeuIleLeuAla 1774

Qy        919 AAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA------------TCTGCTGAC 966
                 ||||||:::      |||   :::::::::::               |||:::
Db       1775 LeuCysGlySerLeuHisTyrGluAsnGluGlnGluArgIleArgGluThrGluAlaGlu 1794

Qy        967 ATACAAAATCAT-----------------------------ACATTAGAGACAGTA--- 993
              :::|||                                 ||||||:::|||:::
Db       1795 ValGlnProGluValValThrThrProGluProValTyrIleThrLeuAspThrLeuAla 1814

Qy        994 --------------------------AATATTTCTGATGTTAATGATTTTCAAATA--- 1023
                                       |||   |||       :::   :::
Db       1815 LeuGlnLysSerThrSerThrSerAsnAsnSerSerGlySerGluGluIleValValMet 1834

Qy       1024 ----------------------------AGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
                                          |||         ::::::      :::
Db       1835 GluGluProValLysLysLeuSerLysLysHisLysArgLeuLysGlnLeuGlnLeuGln 1854

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
                          :::   |||   ||||||||:::|||
Db       1855 GlnGlnGlnGlnGlnArgGluArgGluProAspAspGluGluLeuArgProLeuIleSer 1874

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173
              |||         |||:::      |||      :::|||   :::
Db       1875 IleSerMetCysAspSerGlnLeuAspValProAspIleProVal--------------- 1889

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
                          :::   |||   |||||||||:::|||:::
Db       1890 ------------------ArgAlaMetProAspAspAspGluGlyValGluGluGluPro 1903

Qy       1234 GAAGAATTATCTGAAGAAAAAATA---------------------------------- 1257
              |||         |||:::   :::
Db       1904 GluThrGluProGluGlnSerLeuProGluAspLeuLeuHisPheGlySerSerGlyVal 1923

Qy       1258 ----------------------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293
                                    :::|||                |||   |||:::
Db       1924 AlaThrAsnAlaThrThrThrAspSerGluGlyProValProAlaThrThrSerAspAsp 1943

Qy       1294 AATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGAT 1353
              |||                   |||      :::            |||   |||
Db       1944 AsnValValTyrGlySerSerSerThrSerAsnAlaAlaProHisLysLeuLysThrLys 1963

Qy       1354 AAGCAGGTTAATAAG 1368
              |||      :::|||
```

```
Db       1964 LysLeuGluHisLys 1968

RESULT 20
US-09-216-393-81
; Sequence 81, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393-81

Alignment Scores:
Pred. No.:              4.85e-14        Length:         611
Score:                  323.50          Matches:        78
Percent Similarity:     64.66%          Conservative:   83
Best Local Similarity:  31.33%          Mismatches:     69
Query Match:            13.01%          Indels:         19
DB:                     3               Gaps:           9

US-09-837-344-41 (1-1482) x US-09-216-393-81 (1-611)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||      |||:::
Db        323 ArgGluArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||          :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                 |||:::|||||||||                |||   :::   :::|||::::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||       |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::          |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGlu 462

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
              ||||||       |||:::   :::|||          |||:::   :::   |||:::
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495
              |||:::   :::||||||   :::|||:::|||   :::   |||:::|||||||||
Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546
                  :::||||||   :::|||:::|||   :::   ::::::|||||||||
Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606
                 :::||||||   ::::::::::|||   :::       :::||||||:::       |||
Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::|||:::::::|||:::   :::||||||:::   :::   |||       |||
Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690
                 |||   ::::::||||||   :::
Db        563 ValGluLeuGlnArgGluHisAlaGlu 571

RESULT 21
US-10-321-856-81
; Sequence 81, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2-1
;  CURRENT APPLICATION NUMBER: US/10/321,856
;  CURRENT FILING DATE:  2002-12-17
;  PRIOR APPLICATION NUMBER: 09/216,393
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-10-321-856-81

Alignment Scores:
Pred. No.:              4.85e-14        Length:         611
Score:                  323.50          Matches:        78
Percent Similarity:     64.66%          Conservative:   83
Best Local Similarity:  31.33%          Mismatches:     69
Query Match:            13.01%          Indels:         19
DB:                     4               Gaps:           9

US-09-837-344-41 (1-1482) x US-10-321-856-81 (1-611)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTG 51
              :::|||:::|||   :::   |||:::|||||||||      |||:::
Db        323 ArgGluArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArg 342

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA---CAAGAG 108
              |||   ::::::   :::   |||:::   |||:::   :::||||||      |||
Db        343 GlnArgGluGluAspGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArg 362

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              ::::::   :::   |||:::   |||:::   :::||||||   ::::::::::::
Db        363 GluGluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGlu 382

Qy        169 GATCTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
              :::   |||:::|||||||||          :::||||||   ::::::::::::   :::
Db        383 GluGluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGlu 402

Qy        223 TTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
                 |||:::|||||||||                |||   :::   :::|||::::::
Db        403 GluGluArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGlu 422

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
              :::   |||       |||:::   :::||||||   ::::::::::::   :::   |||
Db        423 GluArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGlu 442

Qy        331 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::|||   ||||||   :::|||:::          |||:::   :::   |||:::|||
Db        443 ArgGluGlyArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGlu 462

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAA---------CAACAAAGCGATTTAGAACAAGAG 438
```

```
Db        463 ArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluArgGluArgArg 482

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 495

Db        483 ArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArgGluArgArgArg 502

Qy        496 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------ 546

Db        503 ValGluGluGluLysGluArgGluArgGlnGluGluGluLysArgGluArgArgArgVal 522

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 606

Db        523 GluGluGluLysAlaArgGlnArgGlnGluGluGluGlyArgGluArgGlnArgGlyGlu 542

Qy        607 AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666

Db        543 GluArgGluGluArgGluArgGluPheGlnGlnArgGluArgGluLeuLysThrArgLeu 562

Qy        667 ---GAAAGAAAAAAGGAACATGGAGAT 690

Db        563 ValGluLeuGlnArgGluHisAlaGlu 571

RESULT 22
US-11-097-143-23688
; Sequence 23688, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23688
;   LENGTH: 658
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-23688

Alignment Scores:
Pred. No.:                8.4e-14        Length:          658
Score:                    320.00         Matches:         119
Percent Similarity:       47.22%         Conservative:    119
Best Local Similarity:    23.61%         Mismatches:      202
Query Match:              12.87%         Indels:          64
DB:                       6              Gaps:            18

US-09-837-344-41 (1-1482) x US-11-097-143-23688 (1-658)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69

Db        186 GlnArgGlnArgGluArgGluArgArgArgGlnLeuArgArgGlnGluArgGln------ 203

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129

Db        204 -----------LysLeuAlaArgGluGlnLysArgGluArgGlnArgArgLeuLysGlu 219

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189

Db        220 GluArg---------GlnArgLeuGlnArgGlu-----------GluGlnGlnArgArg 232

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------------- 225

Db        233 GlnArgLeuHisHisAspGluProLysProGlnValAsnProGlnValLysProGlnVal 252

Qy        226 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279

Db        253 IleAspGluLeuArgGlnArgSerGlyGluAspLeuAspLysAsnGlnThrAspGluHis 272

Qy        280 CAAGATAGACTTGCTAAAGAA---------AAGTTACAAGAGCAGCAAAGCGATTTA--- 327

Db        273 GluGlnLysLeuArgAsnGluGlnGluLysLysLeuArgGluGluGlnGlnLysGlnArg 292

Qy        328 ---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 384

Db        293 AspGluGlnGluGlnLysAspArgGluGluGlnAspArgLeuLysGlnGluGluGluGln 312

Qy        385 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTT 444

Db        313 AlaArgThrHisGlnLysGluLeuLysGluAsnGlnGluGlnGlnLeuArgGluLeuLys 332

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA--- 501

Db        333 AlaLysGlnGluArgGluLysGlnGluArgAspTyrGlnGlnGlnLysArgGluHisGlu 352

Qy        502 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558

Db        353 LeuGluLeuLeuLysGlnArgGlnAlaGluAlaAspArgGlnHisAlaAlaAspGluGlu 372

Qy        559 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618

Db        373 AlaGluLysLeuArgLeuGluArgIleGlnLysGlnArgGluLeuGluAlaGlnGlnArg 392

Qy        619 CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAG 678

Db        393 ArgGluArgGluGluGlnArgArgLysGlnArgGluGluGlnGluGluGlnAspArgGln 412

Qy        679 GAACATGGAGATATATTAGCAGAG--------------GATTTATATGGT---CGTTTA 720

Db        413 AsnHisAlaLysArgLeuAlaGluGluLysArgLeuHisAspLeuTyrAlaGluArgIle 432

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Db        433 ArgLeuAlaAsnThrGluArgGluLysGlnLeuAlaGluAlaHisGluAlaLysArgLeu 452

Qy        781 TCTTCTTTACCT---CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837

Db        453 GluGluLeuLysLeuGlnGluGlnLeuLysLysGlnGluAspGluArgGlnGluGlnIle 472

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897

Db        473 ArgArgGluGlnGluGluGlu-----------------GluLysArgLeuGluLeuGlu 486

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957

Db        487 Arg------LeuGluGluAlaArg-----------ArgPheGluGluLysGluLeuLys 500

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        501 ArgLeuHisGluGluAsnGlnArgArgGluGluGlnLysLeuGlnArgGluArgGluIle 520

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077

Db        521 AlaLeuArgGluAlaAlaGluLysLysLeuAlaGluGluGluGluMetLeuArgLysGlu 540
```

```
Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA------TATGACAAT 1131
                 ::::::|||               :::   :::   |||         |||
Db        541 ValAlaGluGluGlu--------------ArgLysValLysGlnArgLeuGluAspGlu 555

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAAT 1191
                 :::      |||                 ||||||   |||:::         :::
Db        556 MetArgGlnAlaGluGluAlaArgLysAlaLysGluAlaGluGluArgAlaAlaGluGlu 575

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
                        :::          :::|||   :::   ::::::      ||||||
Db        576 AlaLysAlaAlaGluGlnLysArgArgValGluAlaAlaLysLysLysAlaAspGluGlu 595

Qy       1252 ---------AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAA 1302
                      |||:::::::::   ::::::         ||:::   :::
Db        596 ValLysAlaLysLeuGluGluLysArgArgGluTyrValThrArgIleSerAlaLeuSer 615

Qy       1303 CCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTT 1362
                 |||   |||:::      :::   |||      |||:::   |||      |||:::
Db        616 ProGluAspGlnLysLysPheIleGluMetArgLysArgArgLysGlnLeuLysGlu--- 634

Qy       1363 AATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA--------- 1413
                 ||||||:::::::::      |||   |||      |||
Db        635 LysLysGluArgAspGlnArgAlaLysGluLeuLysArgIleGlnGlnAlaMetArgLeu 654

Qy       1414 ---GACAATGAA 1422
                    |||||||||
Db        655 AspAspAsnGlu 658

RESULT 23
US-11-097-143-41169
; Sequence 41169, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41169
;   LENGTH: 800
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-41169

Alignment Scores:
Pred. No.:              8.37e-14        Length:         800
Score:                  320.00          Matches:        119
Percent Similarity:     47.22%          Conservative:   119
Best Local Similarity:  23.61%          Mismatches:     202
Query Match:            12.87%          Indels:         64
DB:                     6               Gaps:           18

US-09-837-344-41 (1-1482) x US-11-097-143-41169 (1-800)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||         |||:::|||||||||   :::   :::   ||||||:::|||
Db        328 GlnArgGlnArgGluArgGluArgArgArgGlnLeuArgArgGlnGluArgGln------ 345

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                          :::||||||:::|||:::   :::|||:::|||      |||::::::
Db        346 -----------LysLeuAlaArgGluGlnLysArgGluArgGlnArgArgLeuLysGlu 361

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
              ||||||         ::::::||||||   :::             ||||||:::||||||
Db        362 GluArg---------GlnArgLeuGlnArgGlu------------GluGlnGlnArgArg 374

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------------- 225
                 :::          |||   :::   :::
Db        375 GlnArgLeuHisHisAspGluProLysProGlnValAsnProGlnValLysProGlnVal 394

Qy        226 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                    |||      :::|||:::   |||   |||   :::   |||:::|||
Db        395 IleAspGluLeuArgGlnArgSerGlyGluAspLeuAspLysAsnGlnThrAspGluHis 414

Qy        280 CAAGATAGACTTGCTAAAGAA---------AAGTTACAAGAGCAGCAAAGCGATTTA--- 327
                 :::   :::|||      |||         ||||||:::|||:::|||
Db        415 GluGlnLysLeuArgAsnGluGlnGluLysLysLeuArgGluGluGlnGlnLysGlnArg 434

Qy        328 ---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 384
                 |||||||||:::::::   :::|||:::                  :::   :::   ||||||
Db        435 AspGluGlnGluGlnLysAspArgGluGluGlnAspArgLeuLysGlnGluGluGluGln 454

Qy        385 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTT 444
                 |||      :::::::::|||:::|||   |||            :::|||
Db        455 AlaArgThrHisGlnLysGluLeuLysGluAsnGlnGluGlnGlnLeuArgGluLeuLys 474

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA--- 501
                 ||||||::::::   ::::::|||:::   |||   :::|||::::::|||
Db        475 AlaLysGlnGluArgGluLysGlnGluArgAspTyrGlnGlnGlnLysArgGluHisGlu 494

Qy        502 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558
                 |||   |||:::::::::|||::::::   :::::::::      |||   |||:::
Db        495 LeuGluLeuLeuLysGlnArgGlnAlaGluAlaAspArgGlnHisAlaAlaAspGluGlu 514

Qy        559 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
                 ::::::   :::   :::      |||::::::|||      |||      |||
Db        515 AlaGluLysLeuArgLeuGluArgIleGlnLysGlnArgGluLeuGluAlaGlnGlnArg 534

Qy        619 CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAG 678
                 ::::::|||:::   ::::::||||||         ::::::      :::   ::::::
Db        535 ArgGluArgGluGluGlnArgArgLysGlnArgGluGluGlnGluGluGlnAspArgGln 554

Qy        679 GAACATGGAGATATATTAGCAGAG--------------GATTTATATGGT---CGTTTA 720
                 |||         |||||||||                  |||||||||      |||:::
Db        555 AsnHisAlaLysArgLeuAlaGluGluLysArgLeuHisAspLeuTyrAlaGluArgIle 574

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                 :::          |||          :::          :::
Db        575 ArgLeuAlaAsnThrGluArgGluLysGlnLeuAlaGluAlaHisGluAlaLysArgLeu 594

Qy        781 TCTTCTTTACCT---CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
                 |||      |||:::                  |||   ::::::      |||
Db        595 GluGluLeuLysLeuGlnGluGlnLeuLysLysGlnGluAspGluArgGlnGluGlnIle 614

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                 :::          |||                  |||   |||   ::::::
Db        615 ArgArgGluGlnGluGluGlu-----------------GluLysArgLeuGluLeuGlu 628

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                 :::      |||||||||   :::            :::   |||:::||||||   |||
Db        629 Arg------LeuGluGluAlaArg-----------ArgPheGluGluLysGluLeuLys 642

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                       :::|||            |||      :::            :::
Db        643 ArgLeuHisGluGluAsnGlnArgArgGluGluGlnLysLeuGlnArgGluArgGluIle 662
```

```
Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
Db        663 AlaLeuArgGluAlaAlaGluLysLysLeuAlaGluGluGluGluMetLeuArgLysGlu 682

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA------TATGACAAT 1131
Db        683 ValAlaGluGluGlu--------------ArgLysValLysGlnArgLeuGluAspGlu 697

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAAT 1191
Db        698 MetArgGlnAlaGluGluAlaArgLysAlaLysGluAlaGluGluArgAlaAlaGluGlu 717

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
Db        718 AlaLysAlaAlaGluGlnLysArgArgValGluAlaAlaLysLysLysAlaAspGluGlu 737

Qy       1252 ---------AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAA 1302
Db        738 ValLysAlaLysLeuGluGluLysArgArgGluTyrValThrArgIleSerAlaLeuSer 757

Qy       1303 CCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTT 1362
Db        758 ProGluAspGlnLysLysPheIleGluMetArgLysArgArgLysGlnLeuLysGlu--- 776

Qy       1363 AATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA--------- 1413
Db        777 LysLysGluArgAspGlnArgAlaLysGluLeuLysArgIleGlnGlnAlaMetArgLeu 796

Qy       1414 ---GACAATGAA 1422
Db        797 AspAspAsnGlu 800
```

RESULT 24
US-10-028-248A-103
; Sequence 103, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 103
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-248A-103

Alignment Scores:

Pred. No.:	1.05e-13	Length:	1961
Score:	318.50	Matches:	145
Percent Similarity:	44.00%	Conservative:	97
Best Local Similarity:	26.36%	Mismatches:	171
Query Match:	12.81%	Indels:	137
DB:	4	Gaps:	27

US-09-837-344-41 (1-1482) x US-10-028-248A-103 (1-1961)

```
Qy          1 CAAGAAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
```

```
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654

Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714

Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774

Db       1221 ----------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT 834

Db       1232 GluValLysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy        835 ATA-----------------ATAGAAAAAACAAATAGAGAATCTATTACAACA------ 870

Db       1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy        871 -----------------------------------AATGTTGAAGGA---------CGA 885

Db       1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945

Db       1292 SerAspSerLysSerSerLysLeu------ThrLysAspPheSerAlaLeuGluSerGln 1309

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005

Db       1310 LeuGlnAspThrGlnGluLeuGluGlnGluGluAsnArgGlnLysLeuSerLeuSerThr 1329

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065

Db       1330 ------------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGln 1344

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125

Db       1345 LeuGluGluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHis 1364

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185

Db       1365 AlaGlnValAlaAsp--------------MetLysLysLysMetGluAspSerValGly 1379

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245

Db       1380 CysLeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSer 1399

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296

Db       1400 GlnArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeu 1419

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGAT 1353

Db       1420 GlnGlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAla 1437

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383

Db       1438 CysAsnLeuGluLysLysGlnLysLysPhe 1447

RESULT 25
US-10-107-782-103
; Sequence 103, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
;  APPLICANT: Boldog, Ferenc,
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Colman, Steve,
;  APPLICANT:  Edinger, Shlomit,
;  APPLICANT:  Gangolli, Esha,
;  APPLICANT:  Kekuda, Ramesh,
;  APPLICANT:  Li, Li,
;  APPLICANT:  Liu, Xiaohong,
;  APPLICANT:  Malyankar, Uriel,
;  APPLICANT:  Miller, Charles,
;  APPLICANT:  Millet, Isabelle,
;  APPLICANT:  Patturajan, Meera,
;  APPLICANT:  Rothenberg, Mark,
;  APPLICANT:  Sciore, Paul,
;  APPLICANT:  Shenoy, Suresh,
;  APPLICANT:  Shimkets, Richard,
;  APPLICANT:  Si, Jingsheng,
;  APPLICANT:  Smithson, Glennda,
;  APPLICANT:  Spytek, Kimberly,
;  APPLICANT:  Stone, David,
;  APPLICANT:  Taupier, Raymond, jr.,
;  APPLICANT:  Tchernev, Velizar,
;  APPLICANT:  Vernet, Corine,
;  APPLICANT:  Zerhusen, Brian
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 21402-222CIP
;  CURRENT APPLICATION NUMBER: US/10/107,782
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: 10/028,248
;  PRIOR FILING DATE: 2001-12-19
;  PRIOR APPLICATION NUMBER: 60/256,619
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: 60/262,959
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/272,408
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 60/285,189
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: 60/308,039
;  PRIOR FILING DATE: 2001-07-26
;  PRIOR APPLICATION NUMBER: 60/311,266
;  PRIOR FILING DATE: 2001-08-09
;  PRIOR APPLICATION NUMBER: 60/279,344
;  PRIOR FILING DATE: 2001-03-28
;  NUMBER OF SEQ ID NOS: 215
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 103
;   LENGTH: 1961
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-107-782-103

Alignment Scores:
Pred. No.:              1.05e-13      Length:          1961
Score:                  318.50        Matches:         145
Percent Similarity:     44.00%        Conservative:    97
Best Local Similarity:  26.36%        Mismatches:      171
Query Match:            12.81%        Indels:          137
DB:                     4             Gaps:            27

US-09-837-344-41 (1-1482) x US-10-107-782-103 (1-1961)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117

Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150

Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165

Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021
```

```
Qy       166 ----------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                   :::||||||||||:::      |||   :::||||||
Db      1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy       202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
             |||:::   :::||||||:::   ||||||          ||||||:::
Db      1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy       262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                :::||||||      |||:::::|||      :::||||||   |||   :::
Db      1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy       322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
             :::|||:::   |||      :::   :::|||:::||||||          :::   :::
Db      1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy       379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
             |||:::|||                                |||     :::   :::
Db      1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy       406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
             :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db      1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy       466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
             |||:::   ||||||   :::|||      |||   |||   :::|||
Db      1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy       511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      |||||||||
Db      1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy       553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
             :::|||      |||:::::::::                    ||||||   |||
Db      1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy       601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
             |||      :::   ::::::|||               |||||||||   |||
Db      1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy       655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
             |||   |||||||||:::   |||:::
Db      1212 LysAlaAsnLeuGluLysAlaLysGln------------------------------- 1220

Qy       715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                           ||||||||||||||||
Db      1221 --------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy       775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT 834
                     |||   |||      |||   :::           |||::::::
Db      1232 GluValLysValLeuGluGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy       835 ATA-----------------ATAGAAAAAACAAATAGAGAATCTATTACAACA------ 870
                               ::::::          |||   :::   |||
Db      1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy       871 ----------------------------------AATGTTGAAGGA---------CGA 885
                                               ||||||   |||         :::
Db      1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy       886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                |||          |||      ||||||   |||          |||
Db      1292 SerAspSerLysSerSerLysLeu------ThrLysAspPheSerAlaLeuGluSerGln 1309

Qy       946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
             :::|||      :::|||          :::   ::::::::::|||
Db      1310 LeuGlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr 1329

Qy      1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                  ::::::   :::   |||||||||   ::::::   :::   :::
Db      1330 ------------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGln 1344

Qy      1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
                |||   :::|||||||||:::   :::         |||   :::            :::
Db      1345 LeuGluGluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHis 1364

Qy      1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                           |||               :::   |||::::::||||||   :::
Db      1365 AlaGlnValAlaAsp--------------MetLysLysLysMetGluAspSerValGly 1379

Qy      1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                   |||          :::::::::          ::::::|||      |||   ||||||
Db      1380 CysLeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSer 1399

Qy      1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||:::             |||   ||||||||||||
Db      1400 GlnArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeu 1419

Qy      1297 TTTAAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGAT 1353
                :::      |||      |||   |||   :::|||      :::   ::::::
Db      1420 GlnGlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAla 1437

Qy      1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
                     :::   |||:::::::::||||||
Db      1438 CysAsnLeuGluLysLysGlnLysLysPhe 1447
```

RESULT 26
US-11-097-143-5154
; Sequence 5154, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5154
; LENGTH: 554
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-5154

Alignment Scores:

Pred. No.:	1.15e-13	Length:	554
Score:	318.00	Matches:	67
Percent Similarity:	62.70%	Conservative:	86
Best Local Similarity:	27.46%	Mismatches:	69
Query Match:	12.79%	Indels:	22
DB:	6	Gaps:	5

US-09-837-344-41 (1-1482) x US-11-097-143-5154 (1-554)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::          |||:::   ||||||   :::|||:::   :::|||:::
Db        291 GluGluGluLysArgLysGluGluGluArgArgArgGluGluGluArgLysArgGluGlu 310
```

```
Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA------AAGTTACAAGAGCAGCAA 114
Db        311 GluArgLysArgGluGluGluArgLysArgGluGluGluArgLysArgGluGluGluArg 330

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        331 LysArgGluGluGluArgArgLysGluGluGluArgLysLysGluGluGluArgGluArg 350

Qy        175 GAACAAGAGAGA----------------------------------------CGTGCTAAA 195
Db        351 GluGluGluArgLysArgGluHisAsnArgLysLysGluGluGluArgLysArgGluGlu 370

Qy        196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
Db        371 LysArgArgLysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArg 390

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
Db        391 LysGluGluGluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGlu 410

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
Db        411 GluArgArgLysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArg 430

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
Db        431 LysGluGluGluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGlu 449

Qy        436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db        450 GluArgLysAlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArg 469

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
Db        470 ArgGluGluLysArgArgArgGlu------------GluGluGluArgArgLysGluGlu 485

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 615
Db        486 GluArgArgGluGluGluGluLysArgLysGluGluGluArgArgLysAsp------Glu 503

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db        504 GluArgArgArgGluGluGluLysArgLysGluGluGluArgArgGluLysGluArgArg 523

Qy        676 AAGGAACATGGA 687
Db        524 ArgGluGluGly 527
```

RESULT 27
US-10-723-860-58
; Sequence 58, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 1937
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-58

Alignment Scores:

Pred. No.:	1.23e-13	Length:	1937
Score:	317.50	Matches:	156
Percent Similarity:	42.08%	Conservative:	131
Best Local Similarity:	22.87%	Mismatches:	184
Query Match:	12.77%	Indels:	211
DB:	5	Gaps:	35

US-09-837-344-41 (1-1482) x US-10-723-860-58 (1-1937)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT-----------TTAGAACAAGAGAGA------------ 237
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------------- 540
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540
Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249
```

```
Qy       616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db      1250 LysMetCysArgSerLeuGlu----------AspGlnValSerGluLeuLysThrLys 1265

Qy       676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
Db      1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy       727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
Db      1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy       784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA------TCTATA 837
Db      1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy       838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
Db      1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy       889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
Db      1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy       933 ------------------------------------------------------------ 933
Db      1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy       934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
Db      1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy       979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
Db      1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy      1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
Db      1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy      1087 ------------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
Db      1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy      1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA------------------------ 1152
Db      1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy      1153 ------------------------GGAATTTATAAAGAA------------CTAGAAGAT 1176
Db      1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy      1177 TTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGATGAAGGAATAGAA 1224
Db      1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy      1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
Db      1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy      1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
Db      1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy      1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
Db      1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy      1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
Db      1579 ---IleAlaGluLysAspGluGluIle------------AspGlnLeuLysArgAsnHis 1593

Qy      1459 ACTAAA 1464
```

```
Db      1594 ThrArg 1595
```

RESULT 28
US-10-723-860-106
; Sequence 106, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 106
; LENGTH: 1937
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-106

Alignment Scores:

Pred. No.:	1.23e-13	Length:	1937
Score:	317.50	Matches:	156
Percent Similarity:	42.08%	Conservative:	131
Best Local Similarity:	22.87%	Mismatches:	184
Query Match:	12.77%	Indels:	211
DB:	5	Gaps:	35

US-09-837-344-41 (1-1482) x US-10-723-860-106 (1-1937)

```
Qy         4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db       930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy        61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
Db       950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy        94 GAAAAG----------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db       970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy       127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db       990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy       169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db      1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy       202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db      1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy       238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db      1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy       268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
Db      1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy       310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db      1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy       340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
```

```
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447

Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495

Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540

Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540

Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588

Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615

Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675

Db       1250 LysMetCysArgSerLeuGlu------------AspGlnValSerGluLeuLysThrLys 1265

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726

Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783

Db       1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA------TCTATA 837

Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888

Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933

Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy        933 ------------------------------------------------------------ 933

Db       1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy        934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978

Db       1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038

Db       1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086

Db       1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy       1087 -----------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116

Db       1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------------- 1152

Db       1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy       1153 ------------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176

Db       1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy       1177 TTGATAGAGAAAAATGAAAATTTA-----------GATGATTTAGATGAAGGAATAGAA 1224

Db       1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy       1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281

Db       1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy       1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338

Db       1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy       1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398

Db       1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458

Db       1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy       1459 ACTAAA 1464

Db       1594 ThrArg 1595

RESULT 29
US-10-341-434-95
; Sequence 95, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
;   LENGTH: 1170
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-95

Alignment Scores:
Pred. No.:              1.69e-13        Length:         1170
Score:                  315.50          Matches:        115
Percent Similarity:     43.69%          Conservative:   110
Best Local Similarity:  22.33%          Mismatches:     163
Query Match:            12.69%          Indels:         127
DB:                     4               Gaps:           20

US-09-837-344-41 (1-1482) x US-10-341-434-95 (1-1170)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db        235 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 254

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114

Db        255 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 274

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174

Db        275 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 293
```

```
Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db        294 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 313

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
Db        314 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 333

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
Db        334 AspArg--------------------AspArgAspHisAspArgGluArgGluArgGlu 346

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
Db        347 ---ArgGluArgAspArgGluLysGluArgGluArgGluArgGluGluArgGluArgGlu 365

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
Db        366 ArgGluArgGluArgGluArgGluArgGluArgGluArg---------------------- 378

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAA 519
Db        379 -----------------GluArgGluArgAlaArgGluArgAspLysGluArgGluArg 392

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
Db        393 GlnArgAspTrpGluAspLysAspLysGlyArgAspAspArgArgGluLysArgGluGlu 412

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db        413 IleArgGluAspArgAsnProArgAspGlyHisAspGluArgLysSerLysLysArgTyr 432

Qy        640 AGGAAGGCTGATACG---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
Db        433 ArgAsnGluGlySerProSerProArgGlnSerProLysArgArgArgGluHisSer--- 451

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
Db        451 ------------------------------------------------------------ 451

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
Db        452 ----------------------------------------ProAspSerAspAlaTyrAsn 458

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
Db        459 SerGlyAspAspLysAsn-----------GluLysHisArgLeuLeuSerGlnValVal 474

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--- 927
Db        475 ArgProGlnGluSerArgSerLeuSerProSerHisLeuThrGluAspArgGlnGlyArg 494

Qy        928 ------------TCAATAAAACCAGAACAAAAAGAAGAT---------------------- 954
Db        495 TrpLysGluGluAspArgLysProGluArgLysGluSerSerArgArgTyrGluGluGln 514

Qy        955 -----------------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
Db        515 GluLeuLysGluLysValSerSerValAspLysGlnArgGluGlnThrGluIleLeuGlu 534

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
Db        535 SerSerArgMetArgAlaGlnAspIleIleGlyHisHisGlnSerGluAspArgGluThr 554

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
Db        555 SerAspArgAlaHisAspGluAsnLysLysLysAlaLysIleGlnLys---LysProIle 573

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
Db        574 LysLys------------LysLysGluAspAspValGlyIleGluArg------------ 585

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
Db        586 ----------------------------------GlyAsnIleGluThrThrSerGlu 593

Qy       1237 GAA---------TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
Db        594 AspGlyGlnValPheSerProLysLysGlyGlnLysLysLysSerIleGluLysLysArg 613

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
Db        614 LysLysSer------LysGlyAspSerAspIleSerAspGluGluAlaAlaGlnGlnSer 631

Qy       1348 AATGATAAGCAG--------------------GTTAATAAGGAAAAGGAAAAATTCATA 1386
Db        632 LysLysLysArgGlyProArgThrProProIleThrThrLysGluGlu------------ 647

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
Db        648 ------LeuValGluMetCysAsnGlyLysAsnGlyIleLeuGlu 660
```

RESULT 30
US-10-144-198-2
; Sequence 2, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Alignment Scores:

Pred. No.:	1.68e-13	Length:	1564
Score:	315.50	Matches:	115
Percent Similarity:	43.69%	Conservative:	110
Best Local Similarity:	22.33%	Mismatches:	163
Query Match:	12.69%	Indels:	127
DB:	4	Gaps:	20

US-09-837-344-41 (1-1482) x US-10-144-198-2 (1-1564)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        629 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 648

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114
Db        649 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 668

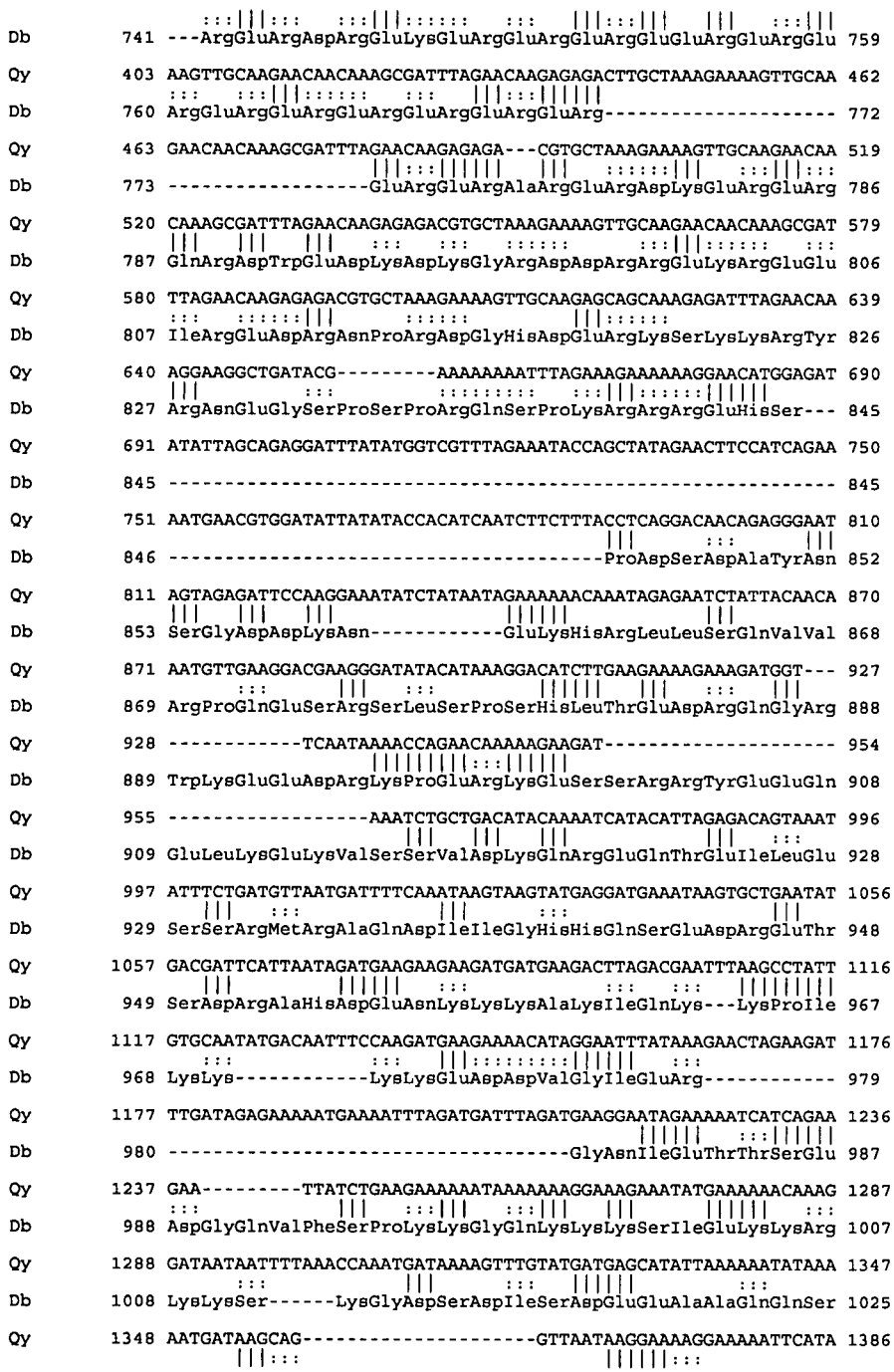
Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        669 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 687

Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db        688 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 707

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
Db        708 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 727

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
Db        728 AspArg--------------------AspArgAspHisAspArgGluArgGluArgGlu 740

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
```



Search completed: December 21, 2005, 11:00:44
Job time : 299.352 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:38:31 ; Search time 8.43173 Seconds
(without alignments)
2507.145 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 53982 seqs, 7132107 residues

Total number of hits satisfying chosen parameters: 107964

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22867/app_query.fasta_1.3740
-DB=Published_Applications_AA_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_2_@runat_21122005_093214_22867
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	315	12.7	1960	7	US-11-069-834-50	Sequence 50, Appl
2	314	12.6	1976	7	US-11-069-834-54	Sequence 54, Appl
3	308	12.4	1976	7	US-11-069-834-52	Sequence 52, Appl
4	304	12.2	1960	7	US-11-069-834-48	Sequence 48, Appl
5	301	12.1	1652	6	US-10-995-561-663	Sequence 663, App
6	301	12.1	1938	6	US-10-995-561-661	Sequence 661, App
7	301	12.1	1938	6	US-10-995-561-662	Sequence 662, App
8	301	12.1	1954	6	US-10-995-561-660	Sequence 660, App
9	301	12.1	1972	6	US-10-995-561-664	Sequence 664, App
10	301	12.1	1972	6	US-10-995-561-666	Sequence 666, App
11	298.5	12.0	1995	7	US-11-069-834-60	Sequence 60, Appl
12	295.5	11.9	1992	7	US-11-069-834-58	Sequence 58, Appl
13	295.5	11.9	2000	7	US-11-069-834-56	Sequence 56, Appl
14	280.5	11.3	1404	6	US-10-878-556A-169	Sequence 169, App
15	275.5	11.1	2101	6	US-10-857-780-23	Sequence 23, Appl
16	266	10.7	793	7	US-11-060-914-2	Sequence 2, Appli
17	262.5	10.6	795	6	US-10-770-726-49	Sequence 49, Appl
18	254	10.2	1388	6	US-10-821-234-1143	Sequence 1143, Ap
19	251.5	10.1	718	6	US-10-878-556A-97	Sequence 97, Appl
20	251	10.1	1266	6	US-10-995-561-665	Sequence 665, App
21	243	9.8	693	7	US-11-196-475-72	Sequence 72, Appl
22	241.5	9.7	885	6	US-10-793-626-1660	Sequence 1660, Ap
23	234.5	9.4	708	7	US-11-196-475-76	Sequence 76, Appl
24	230.5	9.3	1189	7	US-11-074-176-134	Sequence 134, App
25	228	9.2	1586	6	US-10-821-234-901	Sequence 901, App
26	227	9.1	693	7	US-11-196-475-68	Sequence 68, Appl
27	223	9.0	1410	6	US-10-878-556A-136	Sequence 136, App
28	220.5	8.9	989	6	US-10-821-234-975	Sequence 975, App
29	219.5	8.8	700	7	US-11-196-475-74	Sequence 74, Appl
30	216.5	8.7	700	7	US-11-196-475-66	Sequence 66, Appl
31	215.5	8.7	1410	6	US-10-821-234-1050	Sequence 1050, Ap
32	214	8.6	663	7	US-11-196-475-70	Sequence 70, Appl
33	214	8.6	663	7	US-11-196-475-78	Sequence 78, Appl
34	210.5	8.5	558	6	US-10-467-657-694	Sequence 694, App
35	210	8.4	1299	6	US-10-821-234-1145	Sequence 1145, Ap
36	208.5	8.4	898	7	US-11-099-691-7	Sequence 7, Appli
c 37	205	7.8	3500	7	US-11-085-775-2	Sequence 2, Appli
38	199.5	8.0	691	6	US-10-131-826A-16	Sequence 16, Appl
39	199.5	8.0	703	6	US-10-821-234-963	Sequence 963, App
40	199	8.0	744	6	US-10-873-528-184	Sequence 184, App
41	195	7.8	5024	6	US-10-793-626-2964	Sequence 2964, Ap
42	190.5	7.7	664	6	US-10-878-556A-44	Sequence 44, Appl
43	190	7.6	531	7	US-11-060-914-4	Sequence 4, Appli
44	188.5	7.6	693	6	US-10-873-528-185	Sequence 185, App
45	188	7.6	1095	6	US-10-793-626-3154	Sequence 3154, Ap

ALIGNMENTS

RESULT 1
US-11-069-834-50
; Sequence 50, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 50
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-069-834-50

Alignment Scores:

Pred. No.:	1.65e-10	Length:	1960
Score:	315.00	Matches:	145
Percent Similarity:	44.63%	Conservative:	100
Best Local Similarity:	26.41%	Mismatches:	168
Query Match:	12.67%	Indels:	136
DB:	7	Gaps:	28

US-09-837-344-41 (1-1482) x US-11-069-834-50 (1-1960)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ----------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
Db       1212 LysAlaAsnLeuGluLysAlaLysGln------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
Db       1221 --------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
Db       1231 AsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATATCTATAATAGAA--------------------------------AAAACAAAT 852
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG----------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
Db       1311 GlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

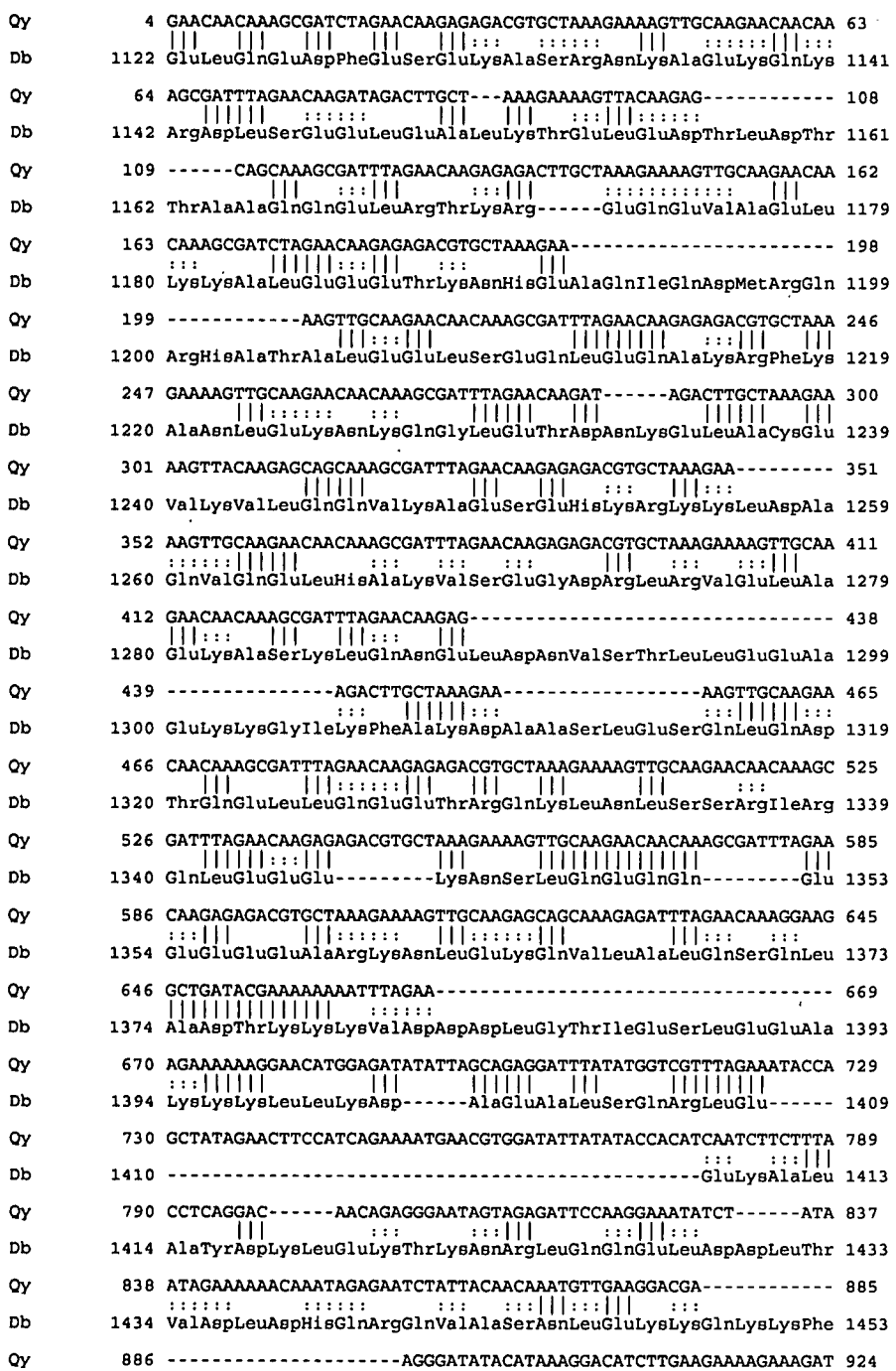
Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446
```

RESULT 2
US-11-069-834-54
; Sequence 54, Application US/11069834
; Publication No. US2005027681lA1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 54
; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-069-834-54

Alignment Scores:

Pred. No.:	1.86e-10	Length:	1976
Score:	314.00	Matches:	146
Percent Similarity:	40.64%	Conservative:	108
Best Local Similarity:	23.36%	Mismatches:	179
Query Match:	12.63%	Indels:	192
DB:	7	Gaps:	26

US-09-837-344-41 (1-1482) x US-11-069-834-54 (1-1976)



RESULT 3
US-11-069-834-52
; Sequence 52, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3

```
; SEQ ID NO 52
;   LENGTH: 1976
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-52

Alignment Scores:
Pred. No.:              3.76e-10        Length:         1976
Score:                  308.00          Matches:        132
Percent Similarity:     44.21%          Conservative:   135
Best Local Similarity:  21.85%          Mismatches:     169
Query Match:            12.39%          Indels:         168
DB:                     7               Gaps:           25

US-09-837-344-41 (1-1482) x US-11-069-834-52 (1-1976)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              :::   :::||||||::      |||      ::::::|||   |||:::::::::
Db        872 GluGlyGluLeuGluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIle 891

Qy         70 TTAGAACAAGAT---------------------------------------------- 81
              |||   :::
Db        892 LeuAlaGluGlnLeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAla 911

Qy         82 AGACTTGCT------AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA------------ 123
              |||||||||      |||:::::|||:::|||         ||||||
Db        912 ArgLeuAlaAlaLysLysGlnGluLeuGluGluIleLeuHisAspLeuGluSerArgVal 931

Qy        124 ------GAACAAGAGAGA---------CTTGCTAAAGAAAAG------------------ 150
                   |||:::||||||         |||      ||||||
Db        932 GluGluGluGluGluArgAsnGlnIleLeuGlnAsnGluLysLysLysMetGlnAlaHis 951

Qy        151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
              :::|||:::   :::      |||::::::|||      |||::::::|||||||||
Db        952 IleGlnAspLeuGluGluGlnLeuAspGluGluGluGlyAlaArgGlnLysLeuGlnLeu 971

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------------TTGCAA 258
              ::::::         |||   :::   :::      |||:::            |||:::
Db        972 GluLysValThrAlaGluAlaLysIleLysLysMetGluGluGluValLeuLeuLeuGlu 991

Qy        259 GAACAACAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              :::|||   |||      :::::::::   :::|||   ::::::::::::   |||
Db        992 AspGlnAsnSerLysPheIleLysGluLysLysLeuMetGluAspArgIleAlaGluCys 1011

Qy        316 CAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA----------------------- 348
                 |||   |||   |||:::|||   :::||||||
Db       1012 SerSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAlaLysIleArgAsnLys 1031

Qy        349 --------------------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
                                        |||:::|||::::::::::          :::
Db       1032 GlnGluValMetIleSerAspLeuGluGluArgLeuLysLysGluGluLysThrArgGln 1051

Qy        382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
              :::      :::||||||   ||||||      :::   :::||||||   |||::::::
Db       1052 GluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAspLeu---GlnAspGln 1070

Qy        442 CTTGCTAAA---------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 486
              :::|||:::                :::::::|||:::   |||   :::      |||:::
Db       1071 IleAlaGluLeuGlnAlaGlnValAspGluLeuLysValGlnLeuThrLysLysGluGlu 1090

Qy        487 GAG---------------------------------------------------------- 489
              |||
Db       1091 GluLeuGlnGlyAlaLeuAlaArgGlyAspAspGluThrLeuHisLysAsnAsnAlaLeu 1110

Qy        490 ---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
                       |||      :::   ::::::   |||   |||   |||   |||   |||
Db       1111 LysValAlaArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspPheGluSerGlu 1130

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              :::   ::::::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1131 LysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAla 1150

Qy        601 ---AAAGAAAAGTTG----------------------------CAAGAGCAGCAAAGA 627
                 |||   :::|||                            ||||||   :::
Db       1151 LeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThr 1170

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
                 ||||||      |||:::   ||||||   ||||||   :::   ::::::
Db       1171 LysArgGluGlnGluValAlaGluLeuLysLysAlaLeuGluAspGluThrLysAsnHis 1190

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
              :::         :::|||:::      |||                  ||||||
Db       1191 GluAlaGlnIleGlnAspMetArgGlnArgHisAlaThrAlaLeuGluGluLeuSerGlu 1210

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
              :::   |||:::                  :::::::::|||   ::::::   :::|||
Db       1211 GlnLeuGluGlnAlaLysArgPhe------LysAlaAsnLeuGluLysAsnLysGlnGly 1228

Qy        808 AATAGTAGAGATTCCAAGGAA---------------------------------------- 828
                    |||:::||||||
Db       1229 LeuGluThrAspAsnLysGluLeuAlaCysGluValLysValLeuGlnGlnValLysAla 1248

Qy        828 ------------------------------------------------------------ 828

Db       1249 GluSerGluHisLysArgLysLysLeuAspAlaGlnValGlnGluLeuHisAlaLysVal 1268

Qy        829 --------------------ATATCTATAATAGAAAAAACAAATAGA------------ 855
                                  :::   :::   ||||||   |||:::
Db       1269 SerGluGlyAspArgLeuArgValGluLeuAlaGluLysAlaAsnLysLeuGlnAsnGlu 1288

Qy        856 ---GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                 :::::::::::|||   :::|||         :::      |||      ::::::
Db       1289 LeuAspAsnValSerThrLeuLeuGlu---------GluAlaGluLysLysGlyIleLys 1305

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAA 972
              ||||||   :::            |||   :::|||                :::|||
Db       1306 PheAlaLysAspAlaAlaGlyLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeuGln 1325

Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
              |||   :::   :::|||:::|||                :::|||   :::
Db       1326 GluGluThrArgGlnLysLeuAsnLeuSerSer------------ArgIleArgGlnLeu 1341

Qy       1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAA 1092
              |||:::|||                :::||||||   :::::::::|||::::::|||
Db       1342 GluGluGlu---------------LysAsnSerLeuGlnGluGlnGlnGluGluGluGlu 1356

Qy       1093 GACTTAGAC---------GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
              :::                |||   :::   :::                     |||
Db       1357 GluAlaArgLysAsnLeuGluLysGlnValLeuAlaLeuGlnSerGlnLeuAlaAspThr 1376

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203
              :::                |||::::::::::|||   :::         |||:::|||:::
Db       1377 Lys---------------LysLysValAspAspAspLeuGlyThrIleGluSerLeuGlu 1391

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAAAAA 1254
              :::      :::   :::   |||      |||   ||||||         |||||||||
Db       1392 GluAlaLysLysLysLeuLeuLysAspValGluAlaLeuSerGlnArgLeuGluGluLys 1411

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              :::      |||   |||||||||||:::      :::      |||
Db       1412 ValLeuAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAsp 1431

Qy       1315 AGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
                   |||   :::|||   ::::::      ::::::      :::   |||:::
Db       1432 LeuThrValAspLeuAspHisGlnArgGlnIleValSerAsn------LeuGluLysLys 1449

Qy       1372 AAGGAAAAATTC 1383
              ::::::||||||
Db       1450 GlnLysLysPhe 1453

RESULT 4
US-11-069-834-48
; Sequence 48, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
```

```
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 48
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-48

Alignment Scores:
Pred. No.:                6.04e-10        Length:          1960
Score:                    304.00          Matches:         149
Percent Similarity:       44.27%          Conservative:    110
Best Local Similarity:    25.47%          Mismatches:      178
Query Match:              12.23%          Indels:          148
DB:                       7               Gaps:            30

US-09-837-344-41 (1-1482) x US-11-069-834-48 (1-1960)

Qy          1 CAAGAACAA---CAAAGCGATCTAGAA---------------CAAGAGAGACGTGCT--- 39
              |||||||||   |||::::::   |||              :::|||   ||||||
Db        886 GlnGluGlnLeuGlnAlaGluThrGluLeuCysAlaGluAlaGluGluLeuArgAlaArg 905

Qy         40 ------------AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 87
                         |||::::::|||:::|||          |||||||||      |||:::
Db        906 LeuThrAlaLysLysGlnGluLeuGluGluIleCysHisAspLeuGluAla---ArgVal 924

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 147
                 :::|||:::   :::   |||             |||:::::::::::   ::::::
Db        925 GluGluGluGluGluArgCysGlnTyrLeuGlnAlaGluLysLysLysMet---GlnGln 943

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207
                 :::||||||   :::     ||||||::|||     |||::::::|||||||||
Db        944 AsnIleGlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGln 963

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                 ::::::         |||   :::   :::         ||||||::|||   |||
Db        964 LeuGluLysValThrThrGluAlaLysLeuLys---------LysLeuGluGluAspGln 980

Qy        268 AGCGATTTAGAACAAGAT------AGACTTGCTAAAGAAAAG------------------ 303
                 :::|||            :::|||||||||||||||
Db        981 IleIleMetGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArg 1000

Qy        304 ------------------------TTACAAGAGCAGCAA-------------------- 318
                                     :::::::|||::::::
Db       1001 ValAlaGluPheThrThrAsnLeuMetGluGluGluGluLysSerLysSerLeuAlaLys 1020

Qy        319 --------------------------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                                       :::||||||||||:::      |||   :::|||
Db       1021 LeuLysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGlu 1040

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA--------- 402
                 |||          |||:::   :::||||||::   ||||||      |||
Db       1041 Lys---------GlnArgGlnGluLeuGluLysThrArgArgLysLeuGluGlyAspSer 1057

Qy        403 ------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
                                       :::||||||   |||   ::::::|||:::   :::
Db       1058 ThrAspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIleAlaGluLeuLysMetGln 1077

Qy        439 AGACTTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--- 489
                 |||||||||      |||:::||||||         :::   :::|||:::|||
Db       1078 ---LeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgValGluGluGluAla 1096

Qy        490 ----------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA 519
                                                                  |||      :::   ::::::   |||
Db       1097 AlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluThrGlnIleSerGluLeu 1116

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||   |||||||||   ||||||   ::::::   |||   ::::::|||:::   |||
Db       1117 GlnGluAspGluGluSerGluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAsp 1136

Qy        580 TTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG---------------------- 612
              |||   :::|||      |||   |||   :::|||
Db       1137 LeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAla 1156

Qy        613 ------CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
                   ||||||   :::         ||||||      :::      ||||||   |||
Db       1157 AlaGlnGlnGluLeuArgSerLysArgGluGlnGluValSerIleLeuLysLysThrLeu 1176

Qy        667 ----------------------------------GAAAGAAAAAAGGAACATGGAGAT 690
                                                |||   :::::::::|||
Db       1177 GluAspGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGln 1196

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAA----------------------ATA 726
                 :::   |||:::|||      :::||||||                      :::
Db       1197 AlaVal---GluGluLeuAlaAspGlnLeuGluGlnThrLysArgValLysAlaThrLeu 1215

Qy        727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786
                       :::        ||||||||||||||                      :::
Db       1216 GluLysAlaLysGlnThrLeuGluAsnGluArgGlyGluLeuAlaAsnGluValLysAla 1235

Qy        787 TTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAGGAAATATCTATA 837
              |||   |||      :::|||:::|||         |||      |||      :::
Db       1236 LeuLeuGln---GlyLysGlyAspSerGluHisLysArgLysLysValGluAlaGlnLeu 1254

Qy        838 ATAGAA---------------------------------AAAACAAATAGAGAATCTATT 864
                 |||                                ::::|||
Db       1255 GlnGluLeuGlnValLysPheSerGluGlyGluArgValArgThrGluLeuAlaAspLys 1274

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAG------ 918
                 |||   ::::::         |||         |||   |||   :::
Db       1275 ValThrLysLeuGlnValGluLeuAspSerValThrGlyLeuLeuSerGlnSerAspSer 1294

Qy        919 ------------------AAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                               ||||||   |||              |||   :::|||
Db       1295 LysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGln 1314

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::|||            :::   :::::::::|||              :::
Db       1315 GluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr------------Lys 1330

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                 :::   :::   |||||||||   ::::::   :::   :::   |||   :::||||||
Db       1331 LeuLysGlnMetGluAspGluLysAsnSer---PheArgGluGlnLeuGluGluGluGlu 1349

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                 |||          :::|||:::      |||   |||
Db       1350 GluAlaLysArgAsnLeuGlu------LysGlnIleAlaThrLeu------------His 1363

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 :::          :::   |||::::::||||||   :::          |||
Db       1364 AlaGlnValThrAspMetLysLysLysMetGluAspGlyValGlyCysLeuGluThrAla 1383

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
                 ::::::               ::::::|||       |||   ||||||         ||||||
Db       1384 GluGluAlaLysArgArgLeuGlnLysAspLeuGluGlyLeuSerGlnArgLeuGluGlu 1403

Qy       1252 AAAATAAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
                 |||:::            |||   ||||||||||||                :::   |||
Db       1404 LysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGlnGlnGluLeuAsp 1423

Qy       1312 AAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                 |||   |||   :::|||      :::   ::::::         :::   |||
Db       1424 AspLeuLeuValAspLeuAspHis------GlnArgGlnSerValSerAsnLeuGluLys 1441

Qy       1369 GAAAAGGAAAAATTC 1383
              :::::::::||||||
```

```
Db       1442 LysGlnLysLysPhe 1446

RESULT 5
US-10-995-561-663
; Sequence 663, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 663
;   LENGTH: 1652
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-663

Alignment Scores:
Pred. No.:              8.83e-10      Length:         1652
Score:                  301.00        Matches:        135
Percent Similarity:     42.78%        Conservative:   108
Best Local Similarity:  23.77%        Mismatches:     171
Query Match:            12.11%        Indels:         154
DB:                     6             Gaps:           26

US-09-837-344-41 (1-1482) x US-10-995-561-663 (1-1652)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::   :::|||:::   |||       ||||||
Db        646 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 665

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::      |||:::::::|||||||||
Db        666 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 685

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                   |||
Db        686 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 705

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::     :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        706 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 725

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db        726 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 745

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   :::::::::::::
Db        746 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 765

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db        766 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 785

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||||::::::
Db        786 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 805

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                 ::::::   |||:::      |||   :::
Db        806 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 825

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                   |||   ||||||:::   ||||||
Db        826 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 845

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db        846 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 865

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::            :::      :::||||||   |||||||||
Db        866 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 885

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::              |||
Db        886 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 905

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db        906 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 925

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::||||||||||      |||   :::   ::::::   ||||||:::
Db        926 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 945

Qy        640 AGGAAGGCTGAT-----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                             |||:::   ::::||
Db        946 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 965

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||:::                                 :::||||||
Db        966 HisLysLysLys-------------------------------LysLeuGluAlaGln 974

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    ||||||                                 ||||||
Db        975 ValGlnGluLeu-----------------------------------GlnSerLysCys 982

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                       |||      :::      :::   ::::::      ::::::
Db        983 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1001

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                    ||||||:::|||                ||||||:::
Db       1002 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1017

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ::::::      ||||||   :::              |||   :::|||
Db       1018 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1035

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::|||      |||   :::   :::|||:::|||                 :::
Db       1036 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1051

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1052 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1070

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::               :::          |||   |||
Db       1071 GluAlaLysGlnAsnLeuGluArgHis---------IleSerThrLeuAsnIleGlnLeu 1087

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::              |||:::|||:::|||                |||   |||
Db       1088 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1104

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::           ||||||
Db       1105 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1124

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||      |||   ||||||||||||:::         :::      |||
Db       1125 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1144

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
                 :::   |||                  |||   :::|||
Db       1145 AspLeuValValAsp--------------LeuAspAsnGlnArgGlnLeuValSerAsn 1159

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||::::::   ||||||
```

```
Db       1160 LeuGluLysLysGlnArgLysPhe 1167

RESULT 6
US-10-995-561-661
; Sequence 661, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 661
;   LENGTH: 1938
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-661

Alignment Scores:
Pred. No.:              8.61e-10        Length:         1938
Score:                  301.00          Matches:        135
Percent Similarity:     42.78%          Conservative:   108
Best Local Similarity:  23.77%          Mismatches:     171
Query Match:            12.11%          Indels:         154
DB:                     6               Gaps:           26

US-09-837-344-41 (1-1482) x US-10-995-561-661 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                            ||||||::::::     |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ------------------------------------AGACTTGCTAAAGAAAAG------ 303
                                                 :::||||||||:::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                 ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::          :::      :::||||||  |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
              :::                            |||  |||:::           |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              ||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                          |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                 :::|||||||
Db       1252 HisLysLysLys--------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                  ||||||                                 ||||||
Db       1261 ValGlnGluLeu-----------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                       |||      :::      :::   ::::::     ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||:::|||                ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ::::::      ||||||   :::          |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              :::|||      |||   :::   :::|||:::|||          :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr-----------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::                  :::         |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis----------IleSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::          |||:::|||:::|||          |||   |||
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::         ||||||
Db       1391 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||               |||   ||||||||||||:::        :::      |||
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
              :::   |||                      |||   :::|||
Db       1431 AspLeuValValAsp--------------LeuAspAsnGlnArgGlnLeuValSerAsn 1445

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
              :::   |||::::::   ||||||
```

```
Db       1446 LeuGluLysLysGlnArgLysPhe 1453

RESULT 7
US-10-995-561-662
; Sequence 662, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 662
;   LENGTH: 1938
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-662

Alignment Scores:
Pred. No.:              8.61e-10      Length:          1938
Score:                  301.00        Matches:         135
Percent Similarity:     42.78%        Conservative:    108
Best Local Similarity:  23.77%        Mismatches:      171
Query Match:            12.11%        Indels:          154
DB:                     6             Gaps:            26

US-09-837-344-41 (1-1482) x US-10-995-561-662 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::     |||:::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-----------------------------------------------AGCGAT 120
              ::::::                                               |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::     |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                 ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                   |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::            :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::           |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                           |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||:::                                 :::||||||
Db       1252 HisLysLysLys-------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    ||||||                                 ||||||
Db       1261 ValGlnGluLeu-----------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                       |||      :::      :::   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                    ||||||:::|||                ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ::::::      ||||||   :::             |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                    :::|||      |||   :::   :::|||:::|||                :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      :::::::|||:::               :::          |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------IleSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::            |||:::|||:::|||                |||   |||
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::           ||||||
Db       1391 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||                |||   ||||||||||||:::         :::      |||
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
              :::   |||                     |||   :::|||
Db       1431 AspLeuValValAsp---------------LeuAspAsnGlnArgGlnLeuValSerAsn 1445

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
              :::   |||::::::   ||||||
```

```
Db       1446 LeuGluLysLysGlnArgLysPhe 1453

RESULT 8
US-10-995-561-660
; Sequence 660, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 660
;   LENGTH: 1954
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-660

Alignment Scores:
Pred. No.:              8.6e-10       Length:          1954
Score:                  301.00        Matches:         135
Percent Similarity:     42.78%        Conservative:    108
Best Local Similarity:  23.77%        Mismatches:      171
Query Match:            12.11%        Indels:          154
DB:                     6             Gaps:            26

US-09-837-344-41 (1-1482) x US-10-995-561-660 (1-1954)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||:::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ----------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                       |||                 ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::             :::      :::||||||   ||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                         |||   |||:::                |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              ||||||||   :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                          |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                 :::||||||
Db       1252 HisLysLysLys--------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                  ||||||                                 ||||||
Db       1261 ValGlnGluLeu-----------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                        |||      :::      :::   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                  ||||||:::|||                ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                  ::::::      ||||||   :::          |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                  :::|||      |||   :::   :::|||:::|||          :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::                :::          |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------IleSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                  ::::::          |||:::|||:::|||          |||   |||
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::          ||||||
Db       1391 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||                |||   ||||||||||||:::          :::      |||
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
                  :::   |||                      |||   :::|||
Db       1431 AspLeuValValAsp--------------LeuAspAsnGlnArgGlnLeuValSerAsn 1445

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
                  :::   |||::::::   ||||||
```

```
Db       1446 LeuGluLysLysGlnArgLysPhe 1453

RESULT 9
US-10-995-561-664
; Sequence 664, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 664
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-664

Alignment Scores:
Pred. No.:              8.58e-10      Length:          1972
Score:                  301.00        Matches:         135
Percent Similarity:     42.78%        Conservative:    108
Best Local Similarity:  23.77%        Mismatches:      171
Query Match:            12.11%        Indels:          154
DB:                     6             Gaps:            26

US-09-837-344-41 (1-1482) x US-10-995-561-664 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                   |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                ::     :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::|||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                  ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::           |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
              ||||||                                 |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||:::                                 :::||||||
Db       1252 HisLysLysLys-------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    ||||||                                 ||||||
Db       1261 ValGlnGluLeu-----------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                       |||      :::      :::   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                    ||||||:::|||                ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ::::::      ||||||   :::             |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::|||      |||   :::   :::|||:::|||                 :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::               :::          |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------IleSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::            |||:::|||:::|||                |||   |||
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::           ||||||
Db       1391 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||                |||   ||||||||||||:::         :::      |||
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
                 :::   |||                     |||   :::|||
Db       1431 AspLeuValValAsp---------------LeuAspAsnGlnArgGlnLeuValSerAsn 1445

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||::::::   ||||||
```

```
Db       1446 LeuGluLysLysGlnArgLysPhe 1453

RESULT 10
US-10-995-561-666
; Sequence 666, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 666
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-666

Alignment Scores:
Pred. No.:                8.58e-10      Length:          1972
Score:                    301.00        Matches:         135
Percent Similarity:       42.78%        Conservative:    108
Best Local Similarity:    23.77%        Mismatches:      171
Query Match:              12.11%        Indels:          154
DB:                       6             Gaps:            26

US-09-837-344-41 (1-1482) x US-10-995-561-666 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                            ||||||::::::     |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::     |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::|||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                  ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                  |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                         |||   |||:::              |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                          |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                  :::||||||
Db       1252 HisLysLysLys--------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||                                     ||||||
Db       1261 ValGlnGluLeu-----------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                       |||      :::      :::   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuAsnAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||:::|||                ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ::::::      ||||||  :::           |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              :::|||      |||  :::   :::|||:::|||          :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||:::|||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||       ::::::|||:::                ::::      |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------IleSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ::::::          |||:::|||:::|||          |||   |||
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrValGluAlaLeu 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::      :::|||      |||   |||:::         ||||||
Db       1391 GluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThrGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||               |||   ||||||||||||:::      :::      |||
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG------------ 1359
              :::   |||                  |||   :::|||
Db       1431 AspLeuValValAsp--------------LeuAspAsnGlnArgGlnLeuValSerAsn 1445

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
              :::   |||::::::   ||||||
```

```
Db       1446 LeuGluLysLysGlnArgLysPhe 1453

RESULT 11
US-11-069-834-60
; Sequence 60, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 60
;   LENGTH: 1995
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-069-834-60

Alignment Scores:
Pred. No.:              1.15e-09        Length:         1995
Score:                  298.50          Matches:        150
Percent Similarity:     39.18%          Conservative:   98
Best Local Similarity:  23.70%          Mismatches:     184
Query Match:            12.01%          Indels:         201
DB:                     7               Gaps:           30

US-09-837-344-41 (1-1482) x US-11-069-834-60 (1-1995)

Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
Db       1127 LeuLysSerLeuArgGluAlaGlnAlaAlaLeuAlaGluAlaGlnGluAspLeuGluSer 1146

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138
Db       1147 GluArgValAlaArgThrLysAlaGluLysGlnArgArgAspLeuGlyGluGluLeuGlu 1166

Qy        139 GCT---------------------------------------------------------- 141
Db       1167 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1186

Qy        142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
Db       1187 SerLysArgGluGlnGluValThrGluLeuLysLysThrLeuGluGluGluThrArgIle 1206

Qy        193 AAAGAA-----------------------------------AAGTTGCAAGAACAACAA 216
Db       1207 HisGluAlaAlaValGlnGluLeuArgGlnArgHisGlyGlnAlaLeuGlyGluLeuAla 1226

Qy        217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
Db       1227 GluGlnLeuGluGlnAlaArgArgGlyLysGlyAlaTrpGluLysThrArgLeuAlaLeu 1246

Qy        277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
Db       1247 GluAlaGluValSerGluLeuArgAlaGluLeuSerSerLeuGlnThrAlaArgGlnGlu 1266

Qy        325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
Db       1267 GlyGluGlnArgArgArgArgLeuGluLeuGlnLeuGlnGluValGlnGlyArgAlaGly 1286

Qy        373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
Db       1287 AspGlyGluArgAlaArgAlaGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1306

Qy        430 GAACAAGAG---------------------------------AGACTTGCTAAAGAA 453

                                                                  
Db       1307 GluAsnValSerGlyAlaLeuAsnGluAlaGluSerLysThrIleArgLeuSerLysGlu 1326

Qy        454 -----------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1327 LeuSerSerThrGluAlaGlnLeuHisAspAlaGlnGluLeuGluGlnGluGluThrArg 1346

Qy        496 GCTAAAGAAAAGTTGCAAGAACAA--------------------------------CAA 522
Db       1347 AlaLysLeuAlaLeuGlySerArgValArgAlaMetGluAlaGluAlaAlaGlyLeuArg 1366

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------------TTGCAAGAACAA 570
Db       1367 GluGlnLeuGluGluGluAlaAlaAlaArgGluArgAlaGlyArgGluLeuGlnThrAla 1386

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA------------------------ 606
Db       1387 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaGlyAlaLeuGluAla 1406

Qy        606 ------------------------------------------------------------ 606
Db       1407 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluAlaLeuThrGlnArgLeuAla 1426

Qy        607 --------------------AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
Db       1427 GluLysThrGluThrValAspArgLeuGluArgGlyArgArgArgLeuGlnGlnGluLeu 1446

Qy        646 GCTGATACGAAAAAAAATTTAGAA--------------------------------AGA 672
Db       1447 AspAspAlaThrMetAspLeuGluGlnGlnArgGlnLeuValSerThrLeuGluLysLys 1466

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
Db       1467 GlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------Ala 1484

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
Db       1485 ValGlu---------GluArgGluArgAla---------------------------- 1491

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
Db       1492 ---GluAlaGluGlyArgGluArgGluAlaArgAlaLeuSerLeuThrArgAlaLeuGlu 1510

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
Db       1511 GluGluGln---------------GluAlaArgGluGluLeu------------------ 1519

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAA------------------CAAAAAGAAGAT 954
Db       1520 GluArgGlnAsnArgAlaLeuArgAlaGluLeuGluAlaLeuLeuSerSerLysAspAsp 1539

Qy        955 AAATCTGCTGACATACAAAATCATACATTAGAG---------ACAGTAAATATTTCTGAT 1005
Db       1540 ValGlyLysSerVal------HisGluLeuGluArgAlaCysArgValAlaGluGlnAla 1557

Qy       1006 GTTAATGATTTT------CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
Db       1558 AlaAsnAspLeuArgAlaGlnValThrGluLeuGluAspGluLeuThrAlaAlaGluAsp 1577

Qy       1060 GATTCATTA----------------------------ATAGATGAAGAAGAAGATGAT 1089
Db       1578 AlaLysLeuArgLeuGluValThrValGlnAlaLeuLysThrGlnHisGluArgAspLeu 1597

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
Db       1598 GlnGlyArgAspGlu--------------------------AlaGlyGluGluArgArg 1608

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTG---ATAGAGAAAAATGAAAATTTAGATGAT 1206
Db       1609 ArgGlnLeuAlaLysGlnLeuArgAspAlaGluValGluArgAspGluGluArgLysGln 1628

Qy       1207 -----------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
```

```
Db       1629 ArgThrLeuAlaValAlaAlaArgLysLysLeuGluGlyGluLeuGluGluLeuLysAla 1648

Qy       1249 GAAAAAATAAAAAAAGGAAAG---AAATATGAAAAAACAAAGGAT---AATAATTTTAAA 1302
              :::           |||:::   |||   |||       |||                :::
Db       1649 GlnMetAlaSerAlaGlyGlnGlyLysGluGluAlaValLysGlnLeuArgLysMetGln 1668

Qy       1303 CCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG--- 1359
                  |||   |||:::   |||   :::::::::   :::      ::::::
Db       1669 AlaGlnMetLysGluLeuTrpArgGlu---ValGluGluThrArgThrSerArgGluGlu 1687

Qy       1360 ---------GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGAC 1410
                     |||:::|||   ||||||   :::|||   |||
Db       1688 IlePheSerGlnAsnArgGluSerGluLysArgLeuLysGlyLeu--------------- 1702

Qy       1411 GGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
                 :::   |||:::|||:::::::   :::||||||:::
Db       1703 ---GluAlaGluValLeuArgLeuGlnGluGluLeuAla 1714

RESULT 12
US-11-069-834-58
; Sequence 58, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 58
;   LENGTH: 1992
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-58

Alignment Scores:
Pred. No.:                1.64e-09        Length:          1992
Score:                    295.50          Matches:         151
Percent Similarity:       39.84%          Conservative:    104
Best Local Similarity:    23.59%          Mismatches:      170
Query Match:              11.89%          Indels:          215
DB:                       7               Gaps:            31

US-09-837-344-41 (1-1482) x US-11-069-834-58 (1-1992)

Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
              |||:::      |||   |||:::      |||   |||   |||   |||||||||:
Db       1123 LeuLysSerLeuArgGluAlaGlnAlaGlyLeuAlaGluAlaGlnGluAspLeuGluAla 1142

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138
              :::|||:::|||:::   |||   :::::::|||:::   ||||||   :::|||
Db       1143 GluArgValAlaArgAlaLysAlaGluLysGlnArgArgAspLeuGlyGluGluLeuGlu 1162

Qy        139 GCT---------------------------------------------------------- 141
              |||
Db       1163 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1182

Qy        142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
                      ::::::::::::   |||   :::      ||||||:::|||   ||||||
Db       1183 SerLysArgGluGlnGluValThrGluLeuLysLysAlaLeuGluGluGluSerArgAla 1202

Qy        193 AAAGAA-----------------------------------AAGTTGCAAGAACAACAA 216
                 |||                                     |||   |||
Db       1203 HisGluValSerMetGlnGluLeuArgGlnArgHisSerGlnAlaLeuValGluMetAla 1222

Qy        217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
                 |||||||||   ||||||   |||         ::::::   :::      |||
Db       1223 GluGlnLeuGluGlnAlaArgArgGlyLysGlyValTrpGluLysThrArgLeuSerLeu 1242

Qy        277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
              |||   :::            |||||   :::      :::   :::
Db       1243 GluAlaGluValSerGluLeuLysAlaGluLeuSerSerLeuGlnThrSerArgGlnGlu 1262

Qy        325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
              ||||||:::||||||   |||   :::|||||||||   |||      |||
Db       1263 GlyGluGlnLysArgArgArgLeuGluSerGlnLeuGlnGluValGlnGlyArgSerSer 1282

Qy        373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
              |||   |||:::   |||      |||   ||||||||||||      |||::::::|||
Db       1283 AspSerGluArgAlaArgSerGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1302

Qy        430 GAACAAGAG---------------------------------AGACTTGCTAAAGAA 453
              |||                                       ||||||   ||||||
Db       1303 GluSerValSerThrAlaLeuSerGluAlaGluSerLysAlaIleArgLeuGlyLysGlu 1322

Qy        454 ------------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
                               :::|||   :::   |||      |||:::::|||   |||
Db       1323 LeuSerSerAlaGluSerGlnLeuHisAspThrGlnGluLeuLeuGlnGluGluThrArg 1342

Qy        496 GCTAAA------------------------------------------------------ 501
              ||||||
Db       1343 AlaLysLeuAlaLeuGlySerArgValArgAlaLeuGluAlaGluAlaAlaGlyLeuArg 1362

Qy        502 ----------------------------------------GAAAAGTTGCAAGAACAA 519
                                                     :::||||||
Db       1363 GluGlnMetGluGluGluValValAlaArgGluArgAlaGlyArgGluLeuGlnSerThr 1382

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 555
              |||:::   |||   :::   ||||||   :::|||
Db       1383 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaAlaValLeuGluAla 1402

Qy        556 ------------------------------------------------AAGTTGCAA 564
                                                             :::|||
Db       1403 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluThrLeuThrGlnArgLeuAla 1422

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG--- 621
              |||:::         :::|||:::      ||||||:::   :::||||||::::::
Db       1423 GluLysThrGluAlaValGluArgLeuGluArgAlaArgArgArgLeuGlnGlnGluLeu 1442

Qy        622 ---------CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
                       ||||||   |||:::|||                  |||||||:::
Db       1443 AspAspAlaThrValAspLeuGlyGlnGlnLys---GlnLeuLeuSerThrLeuGluLys 1461

Qy        673 AAAAAGGAACATGGAGAT---ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              |||:::         |||   :::|||||||||:::            |||
Db       1462 LysGlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------ 1479

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              |||:::|||         :::   ||||||
Db       1480 AlaValGlu---------AspArgGluArg---------------------------- 1486

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                    :::      |||      |||:::::::::   :::|||:::
Db       1487 ---IleGluAlaGluGlyArgGluArgGluAlaArgAlaLeuSerLeuThrArgAlaLeu 1505

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              |||                  |||   |||   ::::::
Db       1506 GluGluGluGln--------------GluAlaArgGluGluLeu--------------- 1515

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAA-----------------CAAAAAGAA 951
                 |||::::::::::   :::::::::   |||                |||:::
Db       1516 ---GluArgGlnAsnArgAlaLeuArgAlaGluLeuGluAlaLeuLeuSerSerLysAsp 1534

Qy        952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT--------------- 996
              |||          ::::::      |||   ||||||
Db       1535 AspValGlyLysAsnVal------HisGluLeuGluArgAlaArgLysAlaAlaGluGln 1552

Qy        997 ---ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
                 ||||||:::         |||:::::::::   |||||||||::::::|||
```

```
Db       1553 AlaAlaSerAspLeuArgThr---GlnValThrGluLeuGluAspGluLeuThrAlaAla 1571

Qy       1054 TATGACGATTCATTAATAGATGAA--------------------GAAGAAGATGATGAA 1092
Db       1572 GluAspAlaLysLeuArgLeuGluValThrValGlnAlaLeuLysAlaGlnHisGluArg 1591

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
Db       1592 AspLeuGln------------------GlyArgAspAspAlaGlyGluGluArgArgArg 1605

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTG---ATAGAGAAAAATGAAAATTTAGATGAT--- 1206
Db       1606 GlnLeuAlaLysGlnLeuArgAspAlaGluValGluArgAspGluGluArgLysGlnArg 1625

Qy       1207 ---------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTA--------- 1242
Db       1626 AlaLeuAlaMetAlaAlaArgLysLysLeuGluLeuGluLeuGluGluLeuLysAlaGln 1645

Qy       1243 --------------------TCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
Db       1646 ThrSerAlaAlaGlyGlnGlyLysGluGluAlaValLysGlnLeuLysLysMetGlnVal 1665

Qy       1282 ACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
Db       1666 GlnMet-----------------------LysGluLeuTrpArgGlu---ValGluGlu 1676

Qy       1342 TATAAAAATGATAAGCAG------------GTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db       1677 ThrArgSerSerArgAspGluMetPheThrLeuSerArgGluAsnGluLysLysLeuLys 1696

Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
Db       1697 GlyLeu------------------GluAlaGluValLeuArgLeuGlnGluGluLeuAla 1710
```

RESULT 13
US-11-069-834-56
; Sequence 56, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 56
; LENGTH: 2000
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-069-834-56

Alignment Scores:

Pred. No.:	1.64e-09	Length:	2000
Score:	295.50	Matches:	151
Percent Similarity:	39.84%	Conservative:	104
Best Local Similarity:	23.59%	Mismatches:	170
Query Match:	11.89%	Indels:	215
DB:	7	Gaps:	31

US-09-837-344-41 (1-1482) x US-11-069-834-56 (1-2000)

```
Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
Db       1131 LeuLysSerLeuArgGluAlaGlnAlaGlyLeuAlaGluAlaGlnGluAspLeuGluAla 1150

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138

Qy        139 GCT---------------------------------------------------------- 141
Db       1171 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1190

Qy        142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
Db       1191 SerLysArgGluGlnGluValThrGluLeuLysLysAlaLeuGluGluGluSerArgAla 1210

Qy        193 AAAGAA-----------------------------------AAGTTGCAAGAACAACAA 216
Db       1211 HisGluValSerMetGlnGluLeuArgGlnArgHisSerGlnAlaLeuValGluMetAla 1230

Qy        217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
Db       1231 GluGlnLeuGluGlnAlaArgArgGlyLysGlyValTrpGluLysThrArgLeuSerLeu 1250

Qy        277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
Db       1251 GluAlaGluValSerGluLeuLysAlaGluLeuSerSerLeuGlnThrSerArgGlnGlu 1270

Qy        325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
Db       1271 GlyGluGlnLysArgArgArgLeuGluSerGlnLeuGlnGluValGlnGlyArgSerSer 1290

Qy        373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
Db       1291 AspSerGluArgAlaArgSerGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1310

Qy        430 GAACAAGAG-----------------------------------AGACTTGCTAAAGAA 453
Db       1311 GluSerValSerThrAlaLeuSerGluAlaGluSerLysAlaIleArgLeuGlyLysGlu 1330

Qy        454 ------------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1331 LeuSerSerAlaGluSerGlnLeuHisAspThrGlnGluLeuLeuGlnGluGluThrArg 1350

Qy        496 GCTAAA----------------------------------------------------- 501
Db       1351 AlaLysLeuAlaLeuGlySerArgValArgAlaLeuGluAlaGluAlaAlaGlyLeuArg 1370

Qy        502 ----------------------------------------GAAAAGTTGCAAGAACAA 519
Db       1371 GluGlnMetGluGluGluValValAlaArgGluArgAlaGlyArgGluLeuGlnSerThr 1390

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA----------------------- 555
Db       1391 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaAlaValLeuGluAla 1410

Qy        556 ---------------------------------------------------AAGTTGCAA 564
Db       1411 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluThrLeuThrGlnArgLeuAla 1430

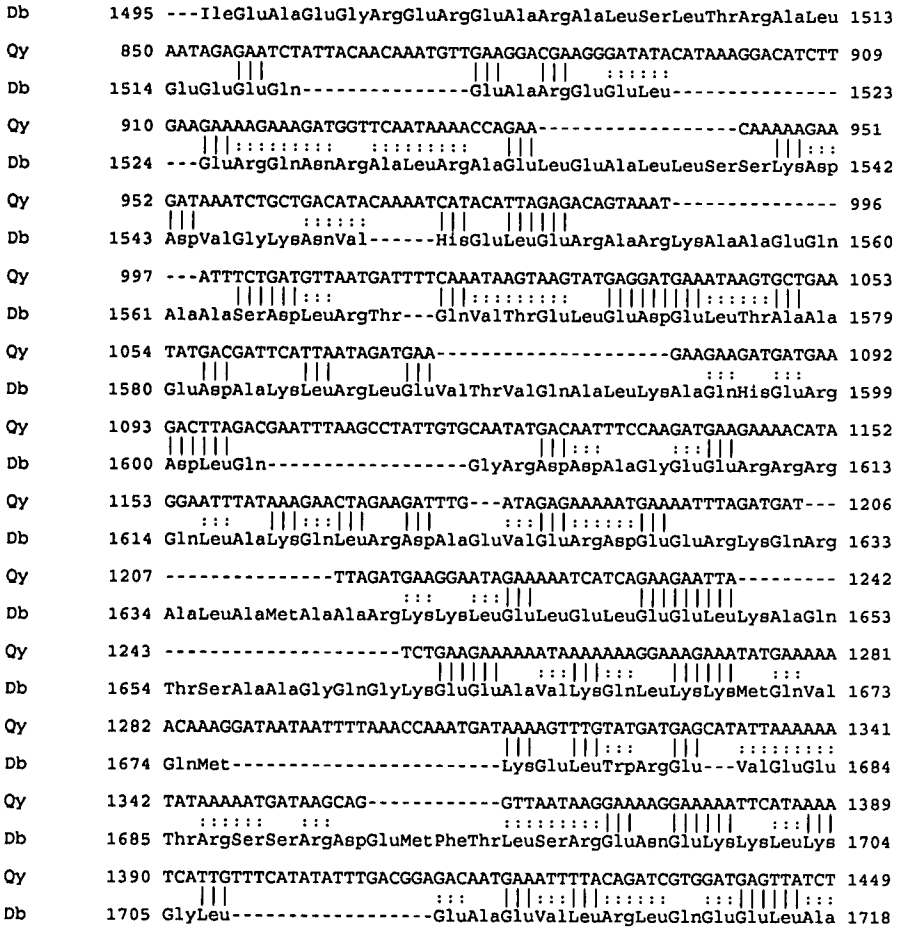
Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG--- 621
Db       1431 GluLysThrGluAlaValGluArgLeuGluArgAlaArgArgArgLeuGlnGlnGluLeu 1450

Qy        622 ---------CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
Db       1451 AspAspAlaThrValAspLeuGlyGlnGlnLys---GlnLeuLeuSerThrLeuGluLys 1469

Qy        673 AAAAAGGAACATGGAGAT---ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db       1470 LysGlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------ 1487

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db       1488 AlaValGlu---------AspArgGluArg---------------------------- 1494

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
```

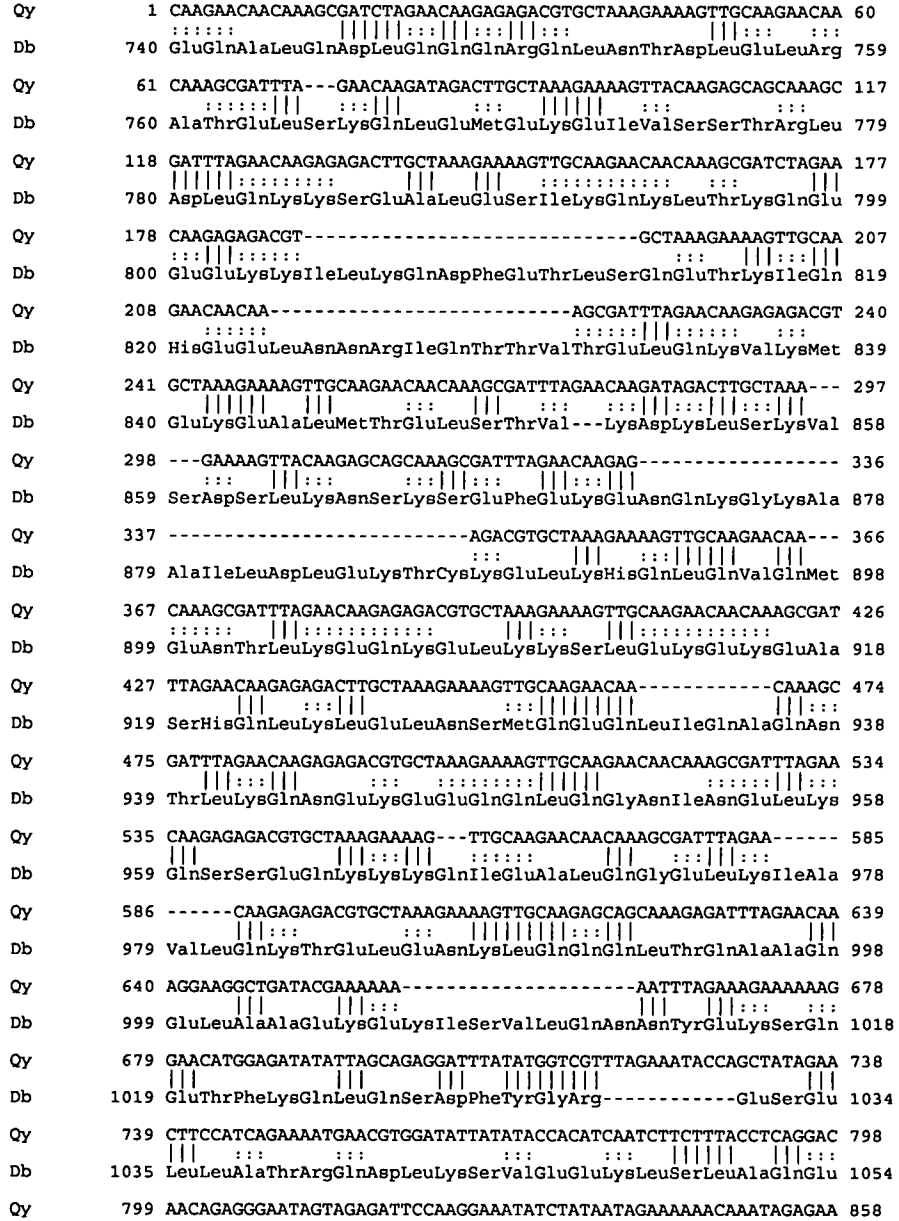


RESULT 14
US-10-878-556A-169
; Sequence 169, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression
; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28
; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 169
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: humangp/chr12-q14221
; DATABASE ENTRY DATE: 2003-04-22
US-10-878-556A-169

Alignment Scores:
Pred. No.: 1.01e-08 Length: 1404

Score:	280.50	Matches: 118
Percent Similarity:	45.74%	Conservative: 140
Best Local Similarity:	20.92%	Mismatches: 207
Query Match:	11.28%	Indels: 99
DB:	6	Gaps: 21

US-09-837-344-41 (1-1482) x US-10-878-556A-169 (1-1404)



```
Db       1055 AspLeuIleSerAsnArgAsn------GlnIleGlyAsnGlnAsnLysLeuIleGlnGlu 1072

Qy        859 TCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA---GAA 915

Db       1073 LeuLysThrAlaLysAlaThrLeuGluGlnAspSerAlaLysLysGluGlnGlnLeuGln 1092

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGAT------AAATCTGCTGACATA 969

Db       1093 GluArgCysLysAlaLeuGlnAspIleGlnLysGluLysSerLeuLysGluLysGluLeu 1112

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029

Db       1113 ValAsnGluLysSerLysLeuAlaGluIleGluGluIleLys------------CysArg 1128

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089

Db       1129 GlnGluLysGluIleThrLys-----------------LeuAsnGluGluLeuLysSer 1142

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149

Db       1143 HisLysLeuGluSerIleLysGluIleThr---------AsnLeuLysAspAlaLysGln 1159

Qy       1150 ATAGGAATTTATAAAGAACTAGAA----------------------------GATTTG 1179

Db       1160 LeuLeuIleGlnGlnLysLeuGluLeuGlnGlyLysAlaAspSerLeuLysAlaAlaVal 1179

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239

Db       1180 GluGlnGluLysArgAsnGlnGlnIleLeuLysAspGlnValLysLysGluGluGluGlu 1199

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAA---------AAAACAAAGGAT 1290

Db       1200 LeuLysLysGluPheIleGluLysGluAlaLysLeuHisSerGluIleLysGluLysGlu 1219

Qy       1291 AATAATTTTAAACCAAATGATAAA--------------------------AGTTTGTAT 1323

Db       1220 ValGlyMetLysLysHisGluGluAsnGluAlaLysLeuThrMetGlnIleThrAlaLeu 1239

Qy       1324 GATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383

Db       1240 AsnGluAsnLeuGlyThrValLysLysGluTrpGlnSerSerGlnArgArgValSerGlu 1259

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAG 1443

Db       1260 LeuGluLysGlnThrAspAspLeuArgGlyGluIleAlaValLeuGluAlaThrValGln 1279

Qy       1444 TTATCTGAAGAT 1455

Db       1280 AsnAsnGlnAsp 1283
```

RESULT 15
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-23

Alignment Scores:

Pred. No.:	1.71e-08	Length:	2101
Score:	275.50	Matches:	133
Percent Similarity:	42.94%	Conservative:	86
Best Local Similarity:	26.08%	Mismatches:	168
Query Match:	11.08%	Indels:	123
DB:	6	Gaps:	24

US-09-837-344-41 (1-1482) x US-10-857-780-23 (1-2101)

```
Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69

Db        472 GlnSerSerIleSerAsnLeuSerGlnAlaLysGluGluLeu---GluGlnAlaSerGln 490

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAA------------------------------- 96

Db        491 AlaHisGlyAlaArgLeuThrAlaGlnValAlaSerLeuThrSerGluLeuThrThrLeu 510

Qy         97 ------AAGTTACAAGAGCAGCAAAGCGATTTAGAA---CAAGAGAGACTTGCTAAAGAA 147

Db        511 AsnAlaThrIleGlnGlnGlnAspGlnGluLeuAlaGlyLeuLysGlnGlnAlaLysGlu 530

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207

Db        531 LysGlnAlaGlnLeuAlaGlnThrLeuGlnGlnGlnGluGlnAlaSerGlnGlyLeuArg 550

Qy        208 GAACAA------------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255

Db        551 HisGlnValGluGlnLeuSerSerSerLeuLysGlnLysGluGlnGlnLeuLysGluVal 570

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGA------------------CTTGCTAAA 297

Db        571 AlaGluLysGlnGluAlaThrArgGlnAspHisAlaGlnGlnLeuAlaThrAlaAlaGlu 590

Qy        298 GAAAAG---------TTACAAGAG------------------CAGCAAAGCGATTTAGAA 330

Db        591 GluArgGluAlaSerLeuArgGluArgAspAlaAlaLeuLysGlnLeuGluAlaLeuGlu 610

Qy        331 CAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAACAA---------------------- 366

Db        611 LysGluLysAlaAlaLysLeuGluIleLeuGlnGlnGlnLeuGlnValAlaAsnGluAla 630

Qy        367 ------------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAA------------GAA 402

Db        631 ArgAspSerAlaGlnThrSerValThrGlnAlaGlnArgGluLysAlaGluLeuSerArg 650

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462

Db        651 LysValGluGluLeuGlnAlaCysValGluThrAlaArgGlnGluGlnHisGluAlaGln 670

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522

Db        671 AlaGlnValAlaGluLeuGluLeuGlnLeuArgSer--------------GluGlnGln 685

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAAGAACAACAAAGC 576

Db        686 LysAlaThrGluLysGluArgValAlaGlnGluLysAspGlnLeuGlnGluGlnLeuGln 705

Qy        577 GATTTA----------------------------------GAACAAGAGAGACGTGCT 600

Db        706 AlaLeuLysGluSerLeuLysValThrLysGlySerLeuGluGluGluLysArgArgAla 725

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Db        726 AlaAspAlaLeuGluGluGlnGlnArgCysIleSerGluLeuLysAlaGluThrArgSer 745
```

```
Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--- 717
Db        746 LeuValGluGlnHisLysArgGluArgLysGluLeuGluGluGluArgAlaGlyArgLys 765

Qy        718 ---TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
Db        766 GlyLeuGluAlaArgLeuLeuGlnLeuGlyGlu----------------------Ala 777

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT 834
Db        778 HisGlnAla-----------GluThrGluValLeuArgArgGluLeuAlaGluAlaMet 793

Qy        835 ATAATAGAAAAAACAAATAGA---------GAATCTATTACAACAAATGTTGAAGGACGA 885
Db        794 AlaAlaGlnHisThrAlaGluSerGluCysGluGlnLeuValLysGluValAlaAlaTrp 813

Qy        886 AGGGATATACATAAA---GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAA 942
Db        814 ArgAspGlyTyrGluAspSerGlnGlnGluGluAlaGlnTyrGlyAlaMetPheGlnGlu 833

Qy        943 CAAAAA--------------------GAAGATAAATCTGCTGACATACAAAATCATACA 981
Db        834 GlnLeuMetThrLeuLysGluGluCysGluLysAlaArgGlnGluLeuGlnGluAlaLys 853

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT--- 1038
Db        854 GluLysValAlaGlyIleGluSerHisSerGluLeuGlnIleSerArgGlnGlnAsnLys 873

Qy       1039 ------GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT--- 1089
Db        874 LeuAlaGluLeuHisAlaAsnLeuAlaArgAlaLeuGlnGlnValGlnGluLysGluVal 893

Qy       1090 -----------------GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
Db        894 ArgAlaGlnLysLeuAlaAspAspLeuSerThrLeuGln-------------------- 906

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA---------CTAGAAGATTTGATA 1182
Db        907 ---------GluLysMetAlaAlaThrSerLysGluValAlaArgLeuGluThrLeuVal 923

Qy       1183 GAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTA 1242
Db        924 ArgLys-----------------AlaGlyGluGlnGlnGluThrAlaSerArgGluLeu 937

Qy       1243 TCTGAAGAAAAAATAAAAAAAGGAAAGAAA 1272
Db        938 ValLysGluProAlaArgAlaGlyAspArg 947
```

RESULT 16
US-11-060-914-2
; Sequence 2, Application US/11060914
; Publication No. US20050261184A1
; GENERAL INFORMATION:
; APPLICANT: Kaufman, Paul L
; APPLICANT: Geiger, Benjamin
; APPLICANT: Bershadsky, Alexander
; APPLICANT: Borras, Teresa
; TITLE OF INVENTION: Method for Treating Glaucoma
; FILE REFERENCE: 960296.00128
; CURRENT APPLICATION NUMBER: US/11/060,914
; CURRENT FILING DATE: 2005-02-18
; PRIOR APPLICATION NUMBER: US 60/545,722
; PRIOR FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: US 60/545,723
; PRIOR FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-060-914-2

Alignment Scores:

Pred. No.:	6.1e-08	Length:	793
Score:	266.00	Matches:	120
Percent Similarity:	43.55%	Conservative:	103
Best Local Similarity:	23.44%	Mismatches:	199
Query Match:	10.70%	Indels:	90
DB:	7	Gaps:	20

US-09-837-344-41 (1-1482) x US-11-060-914-2 (1-793)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
Db        182 GluAsnLysLysGluAspLysGluLysGluGluGluGluGluLysProLysArgGly 201

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
Db        202 SerIleGlyGluAsnGlnValGluVal-----------MetValGluGluLysThrThr 217

Qy        106 GAGCAGCAAAGCGAT----------------------------TTAGAACAAGAGAGA 135
Db        218 GluSerGlnGluGluThrValValMetSerLeuLysAsnGlyGlnIleSerSerGluGlu 237

Qy        136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------------CTAGAACAA 180
Db        238 ProLysGlnGluGluGluArgGluGlnGlySerAspGluIleSerHisHisGluLysMet 257

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        258 GluGluGluAspLysGluArgAlaGluAlaGluArgAlaArgLeuGluAlaGluGluArg 277

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db        278 ------GluArgIleLysAlaGluGlnAspLysLysIleAlaAspGluArgAlaArgIle 295

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db        296 GluAlaGluGluLysAlaAlaAlaGlnGluArgGluArgArgGluAlaGlu--------- 312

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db        313 GluArgGluArgMetArgGluGluGluLysArgAlaAlaGlu---------GluArgGln 329

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        330 ArgIleLysGluGluGluLysArgAlaAlaGlu---------GluArgGlnArgIleLys 346

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        347 GluGluGluLysArgAlaAlaGlu---------GluArgGlnArgIleLysGluGluGlu 363

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        364 LysArgAlaAlaGlu------GluArgGlnArgAlaArgAlaGluGluGluGluLysAla 381

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
Db        382 LysValGluGluGlnLysArgAsnLysGlnLeuGluGluLys-----------LysHis 397

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT------ 714
Db        398 AlaMetGlnGluThrLysIleLysGlyGluLysValGluGlnLysIleGluGlyLysTrp 417

Qy        715 -------------------------CGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
Db        418 ValAsnGluLysLysAlaGlnGluAspLysLeuGlnThrAlaValLeuLysLysGlnGly 437

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
Db        438 Glu---GluLysGlyThrLysVal--------------------GlnAlaLysArgGlu 449

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
```

```
Db        450 LysLeuGlnGluAspLysProThrPheLysLysGluGluIleLysAspGluLysIleLys 469

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927

Db        470 LysAspLysGluProLysGluGluVal---LysSerPheMetAspArgLysLysGlyPhe 488

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA------ 981

Db        489 ThrGluValLysSerGlnAsnGlyGluPheMetThrHisLysLeuLysHisThrGluAsn 508

Qy        982 ----------------------------TTAGAGACAGTAAATATTTCTGATGTTAAT 1011

Db        509 ThrPheSerArgProGlyGlyArgAlaSerValAspThrLysGluAlaGluGlyAlaPro 528

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATA----------------------AGT 1047

Db        529 GlnValGluAlaGlyLysArgLeuGluGluLeuArgArgArgArgGlyGluThrGluSer 548

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107

Db        549 GluGluPheGluLysLeuLysGlnLysGlnGlnGluAlaAlaLeuGluLeuGluGluLeu 568

Qy       1108 AAGCCTATTGTGCAATATGACAAT------TTCCAAGATGAAGAAAACATAGGAATTTAT 1161

Db        569 LysLysLysArgGluGluArgArgLysValLeuGluGluGluGluGlnArgArgLysGln 588

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221

Db        589 GluGluAlaAspArgLysLeuArgGluGluGluGluLysArgArgLeuLysGluGluIle 608

Qy       1222 GAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA---------AAAGGAAAGAAA 1272

Db        609 GluArgArgArgAlaGluAlaAlaGluLysArgGlnLysMetProGluAspGlyLeuSer 628

Qy       1273 TATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332

Db        629 AspAspLysLysProPheLysCysPheThrProLysGlySerSerLeuLysIleGluGlu 648

Qy       1333 ATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368

Db        649 ArgAlaGluPheLeuAsn---LysSerValGlnLys 659

RESULT 17
US-10-770-726-49
; Sequence 49, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Brown, Eugene
;  APPLICANT:  Liu, Wei
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION:  CANCERS
;  FILE REFERENCE: AM101079 (031896-010000)
;  CURRENT APPLICATION NUMBER: US/10/770,726
;  CURRENT FILING DATE:  2004-02-04
;  NUMBER OF SEQ ID NOS: 48640
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
;   LENGTH: 795
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-770-726-49

Alignment Scores:
Pred. No.:              9.21e-08        Length:         795
Score:                  262.50          Matches:        105
Percent Similarity:     42.47%          Conservative:   112
Best Local Similarity:  20.55%          Mismatches:     177
Query Match:            10.56%          Indels:         117
DB:                     6               Gaps:           17

US-09-837-344-41 (1-1482) x US-10-770-726-49 (1-795)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db         18 GlnGluLysLysArgArgLysGluGlnGluGluLysAlaGluIleLysArgLeuLysAsn 37

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Db         38 SerAspAspArgAspSerLysArg---------AspSerLeuGluGlu------GlyGlu 52

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Db         53 LeuArgAspHisArgMet---------GluIleThrIleArgAsnSerProTyrArgArg 69

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------GATTTAGAACAA 231

Db         70 GluAspSerMetGluAspArgGlyGluGluAspAspSerLeuAlaIleLysProProGln 89

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291

Db         90 GlnMetSerArgLysGluLysValHisHisArg--------------LysAspGluLys 104

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351

Db        105 ArgLysGluLysArgArgHisArgSerHisSerAlaGluGlyGlyLysHisAlaArgVal 124

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411

Db        125 LysGluLysGluArg-----------GluHisGluArgArgLysArgHisArgGluGlu 140

Qy        412 GAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 471

Db        141 GlnAspLysAlaArgArgGluTrpGluArgGlnLysArgArgGluMetAlaArgGluHis 160

Qy        472 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531

Db        161 Ser---------------ArgArgGluArgAspArgLeuGluGlnLeuGluArgLysArg 175

Qy        532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591

Db        176 GluArgGluArgLysMetArgGluGlnGlnLysGluGlnArgGluGlnLysGluArgGlu 195

Qy        592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGAT 651

Db        196 ArgArgAlaGluGluArgArgLysGluArgGlu---------------------------- 206

Qy        652 ACGAAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711

Db        206 ------------------------------------------------------------ 206

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771

Db        207 AlaArgArgGluValSerAla------------------------------------- 213

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA 831

Db        214 ---HisHisArgThrMetArgGluAspTyrSerAspLysValLysAlaSerHisTrpSer 232

Qy        832 TCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGAT 891

Db        233 ArgSerProProArgProProArgGluArgPheGluLeuGly---AspGlyArgLysPro 251

Qy        892 ATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951

Db        252 GlyGluAlaArgProAlaArgAlaGlnLysProAlaGlnLeuLysGluGluLysMetGlu 271

Qy        952 GATAAATCT------GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005

Db        272 GluArgAspLeuLeuSerAspLeuGln----------------------------Asp 281

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065

Db        282 IleSerAspSerGluArgLysThrSerSerAlaGluSerSerSerAlaGluSerGlySer 301

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
```

```
                 :::|||||||||:::::::|||:::   :::|||   :::
Db        302 GlySerGluGluGluGluGluGluGluGluGluGluGluGlySerThrSerGlu 321

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
              ::::::   :::::::||||||                  |||:::      |||
Db        322 GluSerGluGluGluGluGlu----------------------GluGluGluGluGlu 333

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              :::   |||         :::   :::|||      |||:::|||:::||||||:::|||
Db        334 GluGluGluThrGlySerAsnSerGluGluAlaSerGluGlnSerAlaGluGluValSer 353

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              ||||||::::::   :::   :::      |||:::      :::|||:::
Db        354 GluGluGluMetSerGluAspGlu------GluArgGluAsnGluAsnHisLeuLeuVal 371

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
              :::|||   :::|||                  |||         :::
Db        372 ValProGluSerArgPheAsp--------------------ArgAspSerGlyGluSer 384

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
              :::|||   |||:::                  :::   :::|||      :::
Db        385 GluGluAlaGluGlu-----------------GluValGlyGluGlyThrProGlnSer 398

Qy       1426 TTACAGATCGTGGAT-------------------------------------GAGTTA 1446
                 :::   :::                                        ||||||
Db        399 SerAlaLeuThrGluGlyAspTyrValProAspSerProAlaLeuSerProIleGluLeu 418

Qy       1447 TCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 :::::::::   ||||||      |||
Db        419 LysGlnGluLeuProLysTyrLeuProAlaLeu 429

RESULT 18
US-10-821-234-1143
; Sequence 1143, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1143
;   LENGTH: 1388
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1143

Alignment Scores:
Pred. No.:              2.3e-07       Length:         1388
Score:                  254.00        Matches:        112
Percent Similarity:     42.75%        Conservative:   106
Best Local Similarity:  21.96%        Mismatches:     144
Query Match:            10.22%        Indels:         148
DB:                     6             Gaps:           21

US-09-837-344-41 (1-1482) x US-10-821-234-1143 (1-1388)

Qy          1 CAAGAACAACAAAGCGATCTAGAA------CAAGAGAGACGTGCTAAAGAAAAGTTGCAA 54
              :::|||::::::      ||||||          :::|||   ||||||:::|||:::
Db        582 GluGluArgLysGluArgLeuGluSerLeuAsnIleGlnArgGluLysGluGluLeuGlu 601

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCT---AAAGAAAAGTTACAAGAGCAG 111
              ::::::::::::::|||::::::   |||   |||   :::|||:::|||:::::::::
Db        602 GlnArgGluAlaGluLeuGlnLysValArgLysAlaGluGluGluArgLeuArgGlnGlu 621

Qy        112 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA-------------- 156
                 :::   |||:::||||||:::   :::|||      |||
Db        622 AlaLysGluArgGluLysGluArgIleLeuGlnGluHisGluGlnIleLysLysLysThr 641

Qy        156 ---------------------------------------------------------- 156

Db        642 ValArgGluArgLeuGluGlnIleLysLysThrGluLeuGlyAlaLysAlaPheLysAsp 661

Qy        157 ----------------------------------------------GAACAACAAAGC 168
                                                           :::|||
Db        662 IleAspIleGluAspLeuGluGluLeuAspProAspPheIleMetAlaLysGlnValGlu 681

Qy        169 GATCTAGAACAAGAGAGACGT---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
                 ||||||:::|||::::::      :::|||:::|||:::   |||:::      :::
Db        682 GlnLeuGluLysGluLysLysGluLeuGlnGluArgLeuLysAsnGlnGluLysLysIle 701

Qy        226 GAACAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
              :::         |||||||||      |||::::::          :::|||      |||
Db        702 AspTyrPheGluArgAlaLysArgLeuGluGluIleProLeuIleLysSerAlaTyrGlu 721

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT-------------- 324
              :::   |||:::            |||   |||||||||   :::
Db        722 GluGlnArgIleLysAspMetAspLeuTrpGluGlnGlnGluGluGluArgIleThrThr 741

Qy        325 -----------------------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
                                    ||||||   :::   |||            |||:::
Db        742 MetGlnLeuGluArgGluLysAlaLeuGluHisLysAsnArgMetSerArgMetLeuGlu 761

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ::::::         :::   :::   :::   |||::::::   :::   |||::::::
Db        762 AspArgAspLeuPheValMetArgLeuLysAlaAlaArgGlnSerValTyrGluGluLys 781

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAA---------AAGTTGCAAGAACAACAA 471
                    :::|||||||||||||:::|||          :::|||:::|||::::::
Db        782 LeuLysGlnPheGluGluArgLeuAlaGluGluArgHisAsnArgLeuGluGluArgLys 801

Qy        472 AGCGATTTAGAACAAGAGAGACGTGCT---------AAAGAAAAGTTGCAAGAACAACAA 522
                    ::::::||||||||||             :::||||||   :::|||:::|||
Db        802 ArgGlnArgLysGluGluArgArgIleThrTyrTyrArgGluLysGluGluGluGluGln 821

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
                              |||||||||:::|||::::::   :::::::::   :::
Db        822 ------------------ArgArgAlaGluGluGlnMetLeuLysGluArgGluGluArg 835

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
              |||:::      |||||||||          :::|||:::   |||:::   ::::::|||
Db        836 GluArgAlaGluArgAlaLysArg------GluGluGluLeuArgGluTyrGlnGluArg 853

Qy        643 AAGGCTGATACGAAAAAA------AATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696
                    ||||||          :::||||||||||||          :::
Db        854 ---------ValLysLysLeuGluGluValGluArgLysLysArgGlnArgGlu------ 868

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756
                              |||||||||        |||        |||   |||
Db        869 --------------------LeuGluIleGluGluArgGluArgArgArgGluGluGlu 881

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
              |||         :::         |||         :::|||:::|||         |||
Db        882 Arg------ArgLeuGlyAspSerSerLeuSerArgLysAspSerArgTrpGlyAspArg 899

Qy        817 GATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT 876
              ||||||                  |||   |||   |||:::            :::
Db        900 AspSer-----------------GluGlyThrTrpArgLysGlyProGluAlaAspSer 913

Qy        877 GAAGGACGAAGG--------------GATATACATAAAGGACATCTTGAAGAAAAGAAA 921
              |||   ||||||                  :::      :::|||            ::::::
Db        914 GluTrpArgArgGlyProProGluLysGluTrpArgArgGlyGluGlyArgAspGluAsp 933

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                    :::   |||::::::      :::      |||   ::::::
Db        934 ArgSerHisArgArgAspGluGluArgProArgArgLeuGlyAspAspGluAspArg--- 952

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
                                                                  |||
```

```
Db        953 ------------------------------------------------------------Glu 953

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGAC------ 1095
Db        954 ProSerLeuArgProAspAspAspArgValProArgArgGlyMetAspAspAspArgGly 973

Qy       1096 ----------------------------------------TTAGACGAATTTAAGCCT 1113
Db        974 ProArgArgGlyProGluGluAspArgPheSerArgArgGlyAlaAspAspAspArgPro 993

Qy       1114 ATTGTGCAATATGACAATTTCCAAGAT----------------------GAAGAAAAC 1149
Db        994 ------SerTrpArgAsnThrAspAspAspArgProProArgArgIleAlaAspGluAsp 1011

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
Db       1012 ArgGlyAsnTrpArgHisAlaAspAspAspArgProProArgArgGlyLeuAspGlu--- 1030

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAA 1239
Db       1031 AspArgGlySerTrpArgThrAlaAspGlu 1040

RESULT 19
US-10-878-556A-97
; Sequence 97, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
;   LENGTH: 718
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: humangp/chr7-q9unh0
;   DATABASE ENTRY DATE: 2003-01-07
US-10-878-556A-97

Alignment Scores:
Pred. No.:              3.42e-07      Length:         718
Score:                  251.50        Matches:        116
Percent Similarity:     45.27%        Conservative:   104
Best Local Similarity:  23.87%        Mismatches:     175
Query Match:            10.12%        Indels:         91
DB:                     6             Gaps:           23

US-09-837-344-41 (1-1482) x US-10-878-556A-97 (1-718)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db         44 GlnTyrGlnGlnSerMetGlnLeuGluGluArgAlaGluGlnIleArgSerLysSerHis 63

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db         64 LeuIleGlnValGluArgGluLysMetGlnMetGlu---LeuSerHisLysArgAlaArg 82

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db         83 ValGluLeuGluArgAlaAlaSerThrSerAlaArgAsnTyrGluArgGluValAspArg 102

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        103 AsnGlnGluLeuLeuThrArgIleArgGlnLeuGln---------GluArgGluAlaGly 119

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA------------ 288
Db        120 AlaGluGluLysMetGlnGluGln---------LeuGluArgAsnArgGlnCysGlnGln 136

Qy        289 ---------CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG--- 336
Db        137 AsnLeuAspAlaAlaSerLysArgLeuArgGluLysGluAspSerLeuAlaGlnAlaGly 156

Qy        337 ---------AGACGTGCTAAAGAAAAGTTGCAAGAA---CAACAAAGCGATTTAGAACAA 384
Db        157 GluThrIleAsnAlaLeuLysGlyArgIleSerGluLeuGlnTrpSerValMetAspGln 176

Qy        385 GAGAGACGTGCT-----------------AAAGAAAAGTTGCAAGAACAACAAAGCGAT 426
Db        177 GluMetArgValLysArgLeuGluSerGluLysGlnGluLeuGlnGlu---GlnLeuAsp 195

Qy        427 TTAGAACAAGAGAGACTT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        196 LeuGlnHisLysLysCysGlnGluAlaAsnGlnLysIleGlnGluLeuGlnAla------ 213

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        214 SerGlnGluAlaArgAlaAspHis------GluGlnGlnIleLysAspLeuGluGln--- 230

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        231 ---------LysLeuSerLeuGlnGluGlnAspAlaAlaIleValLysAsnMetLysSer 247

Qy        601 AAA---GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db        248 GluLeuValArgLeuProArgLeuGluArgGluLeuLysGlnLeuArgGluGluSer--- 266

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT--- 714
Db        267 AlaHisLeuArgGluMetArgGluThrAsnGlyLeuLeuGlnGluGluLeuGluGlyLeu 286

Qy        715 ------------------------CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
Db        287 GlnArgLysLeuGlyArgGlnGluLysMetGlnGluThrLeuValGlyLeuGluLeuGlu 306

Qy        751 AATGAACGTGGATAT---TATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
Db        307 AsnGluArgLeuLeuAlaLysLeuGlnSerTrpGluArgLeuAspGlnThrMetGlyLeu 326

Qy        808 AATAGTAGAGATTCCAAGGAAATATCT------ATAATAGAAAAAACAAATAGAGAA--- 858
Db        327 SerIleArgThrProGluAspLeuSerArgPheValValGluLeuGlnGlnArgGluLeu 346

Qy        859 ---------------------TCTATTACAACAAATGTTGAAGGA--------------- 882
Db        347 AlaLeuLysAspLysAsnSerAlaValThrSerSerAlaArgGlyLeuGluLysAlaArg 366

Qy        883 -----------------CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db        367 GlnGlnLeuGlnGluGluLeuArgGlnValSerGlyGlnLeuLeuGluGluArgLys--- 385

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
Db        386 ---------LysArgGluThrHisGluAlaLeuAlaArgArgLeuGlnLysArgValLeu 402

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
Db        403 LeuLeuThrLysGluArgAspGlyMetArgAlaIleLeuGlySerTyrAspSerGluLeu 422

Qy       1045 AGT---GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
Db        423 ThrProAlaGluTyrSerProGlnLeuThrArgArgMetArgGluAlaGluAspMet--- 441

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        442 ------------ValGlnLysValHisSerHisSerAlaGluMetGluAlaGlnLeuSer 457

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221
Db        458 GlnAlaLeuGluGluLeu-----------------------------------GlyGly 465

Qy       1222 GAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
```

```
            :::|||         ::::::   |||   ::::::|||      |||
Db        466 GlnLysGlnArgAlaAspMetLeuGluMetGluLeuLysMetLeuLysSerGlnSerSer 485

Qy       1282 ACAAAGGATAATAATTTT 1299
              :::   :::   :::|||
Db        486 SerAlaGluGlnSerPhe 491

RESULT 20
US-10-995-561-665
; Sequence 665, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 665
;   LENGTH: 1266
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-665

Alignment Scores:
Pred. No.:                3.32e-07      Length:          1266
Score:                    251.00        Matches:         114
Percent Similarity:       42.18%        Conservative:    91
Best Local Similarity:    23.46%        Mismatches:      175
Query Match:              10.10%        Indels:          106
DB:                       6             Gaps:            19

US-09-837-344-41 (1-1482) x US-10-995-561-665 (1-1266)

Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
              |||   |||      |||   :::|||::::::|||   ::::::   :::|||::::::
Db        844 LeuLeuGlnValThrArgGlnGluGluGluMetGlnAlaLysGluAspGluLeuGlnLys 863

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG------ 132
                 :::      ::::::|||   :::   :::      :::|||||||||:::
Db        864 ThrLysGluArgGlnGlnLysAlaGluAsnGluLeuLysGluLeuGluGlnLysHisSer 883

Qy        133 AGACTTGCTAAAGAAAAG------TTGCAAGAACAA---CAAAGCGATCTAGAA------ 177
              :::|||   :::||||||      ||||||||||||   |||::::::   |||
Db        884 GlnLeuThrGluGluLysAsnLeuLeuGlnGluGlnLeuGlnAlaGluThrGluLeuTyr 903

Qy        178 ----------------------------------CAAGAGAGACGTGCTAAAGAAAAG 201
                                               :::::::::      :::|||
Db        904 AlaGluAlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGluGluIle 923

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
              |||   |||   ::::::   ||||||:::|||            ::::::||||||
Db        924 LeuHisGluMetGluAlaArgLeuGluGluGluGluAspArgGlyGlnGlnLeuGlnAla 943

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
              ::::::      :::   |||      |||          |||:::||||||
Db        944 GluArgLysLysMetAlaGlnGlnMetLeu---------AspLeuGluGluGln------ 958

Qy        322 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
                 ||||||:::|||      |||::::::||||||||   ::::::         |||
Db        959 ---LeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGlu 977

Qy        382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA---------AGCGATTTAGAA 432
                          ||||||   |||      |||   :::            |||   :::
Db        978 ------------AlaLysIleLysLysLeuGluAspGluIleLeuValMetAspAspGln 993

Qy        433 CAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATTTA------ 480
                    :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        994 AsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeuThrThr 1013

Qy        481 ---------GAACAAGAGAGACGTGCTAAAGAA------AAGTTGCAAGAACAACAAAGC 525
                      |||:::|||   :::||||||          ||||||:::   :::
Db       1014 AsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLysHisGlu 1033

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---CAAAGCGATTTA 582
                 :::      |||          :::   |||   :::|||:::   :::   :::|||
Db       1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGlnGluLeu 1053

Qy        583 GAACAAGAGAGACGTGCTAAAGAA------------AAGTTGCAAGAGCAGCAAAGAGAT 630
              |||:::   :::|||      |||                          ||||||      |||
Db       1054 GluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIleAlaAsp 1073

Qy        631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
              |||:::   :::   |||:::   |||       |||   :::|||:::|||
Db       1074 LeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGlu--------- 1090

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
                                                 |||:::                |||      |||
Db       1091 ------------------------GluLeuGlnAlaAlaLeuAlaArgLeuAspAspGlu 1102

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810

Db       1102 ------------------------------------------------------------ 1102

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
                                               |||  :::|||   |||
Db       1103 ------------------------IleAlaGlnLysAsnAsnAlaLeuLysLysIleArg 1114

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930
                 :::||||||              |||          |||:::|||      |||
Db       1115 GluLeuGluGly------------HisIleSerAspLeuGlnGluAspLeuAspSerGlu 1130

Qy        931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
                          |||   :::|||                                |||||
Db       1131 ArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAla 1150

Qy        991 GTA------AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
              :::      :::   |||      |||      ::::::   |||
Db       1151 LeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAla 1170

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                 |||   :::   :::::::::   :::      |||:::|||
Db       1171 LysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHis 1190

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                    :::   :::                  :::            :::   :::
Db       1191 GluAlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGlu 1210

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              :::||||||                  |||||||||        :::   :::|||
Db       1211 GlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGlu 1230

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||   :::   :::|||:::   |||           |||:::      ::::::
Db       1231 LysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluVal 1250

Qy       1285 AAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATAT 1344
                                                        ||||||   ||||||
Db       1251 ------------------------------------------GluHis---LysLysLys 1255

Qy       1345 AAAAATGATAAGCAGGTT 1362
              |||   :::   ||||||
Db       1256 LysLeuGluAlaGlnVal 1261

RESULT 21
US-11-196-475-72
; Sequence 72, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
```

```
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-72

Alignment Scores:
Pred. No.:                9.35e-07      Length:          693
Score:                    243.00        Matches:         96
Percent Similarity:       44.52%        Conservative:    111
Best Local Similarity:    20.65%        Mismatches:      157
Query Match:              9.77%         Indels:          101
DB:                       7             Gaps:            16

US-09-837-344-41 (1-1482) x US-11-196-475-72 (1-693)

Qy          13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
               :::|||:::::::   |||              |||      :::   :::   |||:::
Db         230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy          73 GAACAAGATAGACTTGCTAAA-----------------GAAAAGTTACAAGAGCAGCAA 114
               |||   |||   :::   :::                  |||   ::::::      :::
Db         247 GluLeuAspAsnValHisGluSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy         115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
               ||||||   :::        :::|||      :::|||   :::|||   :::
Db         267 AspGlnLeuGlu---LysAlaThrAspGluGluHisArgLysGluIleGluSerGlnVal 285

Qy         175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
               :::   :::::::::   ||||||:::|||   ::::::      ||||||:::
Db         286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy         235 AGACGTGCTAAAGAAAAGTTG------CAAGAACAACAAAGCGATTTAGAACAAGATAGA 288
                  :::|||::::::||||||          |||          |||:::::::::|||
Db         304 ---LysAlaGlnGlnLysLeuAspPheSerGluAspAsnLeuAspIleGlnArgAspThr 322

Qy         289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
               :::   :::||||||:::||||||      :::::::::   :::|||:::
Db         323 Val---ArgGluLysIleGlnGluAspIleAsnGluIleAsnLysGluLysAsnLeuPro 341

Qy         349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
               :::                     ::::::::::::   :::   ||||||   |||
Db         342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy         409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
               ::::::   |||      |||                                ::::::
Db         362 GluAspLeuGlnGluGlnLeu-------------------------------LysGlu 370

Qy         469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
               ||||||      |||:::|||     ::::::::::::::   ::::::||||||
Db         371 ThrSerAspGluAsnGlnLysArgGluIleGluLysGlnIleGluIleLysLysSerAsp 390

Qy         529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                  |||      :::            |||      :::         ||||||
Db         391 GluGluLeuLeuLysSerLysAspProLysAlaLeuAspLeuAsnArgAspLeuAsnSer 410

Qy         589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
               :::      :::|||||||||::::::   ::::::::::::::   :::   |||:::
Db         411 LysAlaSerSerLysGluLysIleLysGlyLysGluLysGluIleValLysGluLysSer 430

Qy         649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
                                    |||         ||||||:::
Db         431 ----------------------LysValSerLeuGlyAspLeu--------------- 437

Qy         709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                                                        :::|||:::
Db         438 ------------------------------------AspAsnAspGluThrLeuMet 444

Qy         769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC---AAG 825
                  |||            |||   :::|||   :::   :::|||:::   :::   |||
Db         445 ThrProGluAspGlnLysLeuSerGluAspLysLysLeuAspSerLysLysAsnLeuLys 464

Qy         826 GAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA 885
               :::|||   ||||||:::   |||      |||   :::   |||   |||
Db         465 ProValSerGluIleGluArgValAsnGluIleSerLysSerAsnAsnAsnGlu------ 482

Qy         886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                                                        :::
Db         483 ------------------------------------------ValSerLysSerSer 487

Qy         946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
                     ||||||   :::               :::   |||   |||:::      |||
Db         488 ProLeuAspLysProSerTyrSerAspIleAspSerLysGluValValAspAsnLysAsp 507

Qy        1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
               ||||||      :::                        |||         :::
Db         508 ValAsnLeuGlnGluThrLysProGlnAlaLysSerGlnSerThrSerLeuAsnGlnAsp 527

Qy        1066 TTAATAGATGAAGAAGAAGATGATGAAGAC---------TTAGACGAATTTAAGCCTATT 1116
               ||||||          |||      :::        |||:::        ||||||
Db         528 LeuIleThrMetSerIleAspSerSerAsnProValPheLeuGluValIleAspProIle 547

Qy        1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
                                                  |||:::|||:::   :::
Db         548 Thr--------------------------AsnLeuGlyMetLeuGln------------ 554

Qy        1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
                                    |||   ||||||:::   |||:::           :::
Db         555 --------------------LeuIleAspLeuAsnThrGlyValArg------LeuLys 565

Qy        1237 GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
               |||   :::::::::   |||::::::         ||||||:::   |||||| :::::
Db         566 GluSerThrGlnGlnGlyIleGlnArgTyrGlyIleTyrGluArgGluLysAsp-LeuVa 585

Qy        1297 TTTAAACCAAATG------------ATAAAAGTTTGTATGATGAGCATATTAAAAAATAT 1344
               ::::   :::|||              |||:::   :::::::::   |||:::|||::
Db         585 lValIleLysMetAspSerGlyLysAlaLysLeuGlnIleLeuAsnLysLeuGluAsnLe 605

Qy        1345 AAAAATGATAAGC 1357
               :|||:::||||||
Db         605 uLysValIleSer 609

RESULT 22
US-10-793-626-1660
; Sequence 1660, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
;   LENGTH: 885
;   TYPE: PRT
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-1660

Alignment Scores:
Pred. No.:                1.07e-06        Length:         885
Score:                    241.50          Matches:        99
Percent Similarity:       39.53%          Conservative:   101
Best Local Similarity:    19.57%          Mismatches:     147
Query Match:              9.71%           Indels:         159
DB:                       6               Gaps:           17

US-09-837-344-41 (1-1482) x US-10-793-626-1660 (1-885)

Qy          7 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 66
              |||:::      |||  ||||||:::   ||||||:::     :::   :::
Db        489 GlnArgGlnThrAlaGluPheGluArgGlnPheLysGluGlnLysThrGlnAlaGluGln 508

Qy         67 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 126
                ||||||         |||   ::::::             :::      :::
Db        509 LeuSerGluGlnTyrPheSerAlaSerGlnGlnTyrAsnAsnLeuLysGluGlnValHis 528

Qy        127 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGA 186
                  |||      ::::::|||:::   |||::::::   |||:::   |||
Db        529 HisHisGluLeuGluLeuAspArgLeuLysThrGlnGluAlaHisLeuLysAsnGluHis 548

Qy        187 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 246
                |||:::   :::   :::::::::|||      |||      :::::|||
Db        549 ---------GluGluPheGluPheGluLysAsnAspGlyTyrGlnSerAspLysSerLys 565

Qy        247 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA 306
              |||   |||:::|||:::|||:::   |||   :::             :::::::::|||
Db        566 GluThrLeuLysGluLysGlnAsnHisLeuIleGlu---------IleGlnGlnGlnLeu 582

Qy        307 CAAGAGCAGCAAAGCGATTTAGAACAA-------------------------------- 333
              ::::::   :::||||||:::|||:::
Db        583 LysGlnLeuGluSerAspIleGluArgTyrThrGlnLeuSerLysGluGlyLysAlaSer 602

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------- 378
                :::   :::::::::|||   ::::::||||||||||||
Db        603 ThrHisGlnThrGlnGlnGlnLeuHisGlnLysGlnSerAspLeuAlaValValLysGlu 622

Qy        379 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
                  :::   :::::::::      |||:::|||   :::|||   ||||||   |||
Db        623 ArgIleLysSerGlnLysGlnValTyrGluArgLeuAspLysGlnLeuSerAspSerGlu 642

Qy        433 CAAGAGAGACTT---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
              ::::::::::::         ||||||:::::          ||||||
Db        643 ArgGlnLysIleGluValAsnGluLysIleLysLeuPheAsnSerAspGluMetMetGly 662

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
              :::   |||   |||||||||:::||||||         :::::::|||:::
Db        663 LysAspAlaPheGluLysLeuLysGluGln---------IleGlnGlnGlnGluAsnVal 679

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
              ::::::   |||   :::|||   |||::::::::::|||:::|||:::
Db        680 ArgGlnAsnLeuAsnGlnGlnLeuSerGluIleLysGlnGlnArgLys------------ 695

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669

Db        695 ------------------------------------------------------------ 695

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                                          ||||||      ::::::||||||
Db        696 --------------------------------AspLeuAsnGluLysIleGluIle--- 703

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                  ||||||                  ||||||      |||
Db        704 --------------------AsnGluSerGlnLeuGlnLysCysHisGlnAspIleLeu 716

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
```

```
                  :::|||
Db        717 SerIleGluAsn---------------------------------------------- 720

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                                                                  |||
Db        721 ------------------------------------------------------HisTyr 722

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              ::::::   |||                        |||:::   |||       :::
Db        723 GlnAspIleLys--------------------AlaLysGlnSerLysLeuAspValLeu 735

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
                |||||   ::::::   :::|||            :::   :::|||::::::
Db        736 IleAsnHisAlaIleAspHisLeuAsn-----------AspThrTyrGlnLeuThr--- 750

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                  :::           |||||||||                          |||
Db        751 ---ValGluArgAlaArgMetGluTyrAsp----------------------SerAsp 761

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA--- 1146
              |||   :::|||      :::      |||:::                    :::|||
Db        762 GluThrIleAspAsnLeuArgLysLysValLysLeuThrLysMetThrIleAspGluLeu 781

Qy       1147 ---------AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA--------- 1188
                      |||:::         ::::::   |||:::|||   |||:::
Db        782 GlyProValAsnLeuAsnAlaIleGluGlnPheGluGluLeuAsnGluArgTyrThrPhe 801

Qy       1189 ---AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                |||||      |||||  |||     |||      :::   :::
Db        802 LeuAsnGluGlnArgThrAspLeuArgGluAlaLysGluThrLeuGluGlnIleIleHis 821

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
                |||                        :::|||      :::      ||||||
Db        822 Glu-----------------------MetAspLysGluValGluGlyArgPheLysThr 833

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
                  ::::::   |||                ::::||
Db        834 ThrPheHisAlaValGlnAspHisPheThrThrValPheLys----------------- 847

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
                                              :::|||   |||
Db        848 ------------------------------GlnLeuPheGlyGlyGlyGlnAlaGlu 856

Qy       1426 TTACAGATCGTGGATGAG 1443
              |||::::::   ::::::
Db        857 LeuArgLeuThrGluAsp 862

RESULT 23
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 76
;   LENGTH: 708
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Alignment Scores:
Pred. No.:              2.53e-06        Length:         708
Score:                  234.50          Matches:        93
Percent Similarity:     46.27%          Conservative:   118
Best Local Similarity:  20.39%          Mismatches:     161
Query Match:            9.43%           Indels:         84
DB:                     7               Gaps:           19

US-09-837-344-41 (1-1482) x US-11-196-475-76 (1-708)

Qy         49 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA------AAGTTA 102
              :::|||   :::            :::||||||::::::      |||         :::
Db        232 IleGlnGlyGluThrHisLysAlaAspGlnAspLysIleAspThrGluLeuAspAsnIle 251

Qy        103 CAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 162
               |||      |||::::::   :::              |||   |||::::::|||
Db        252 HisGluSerAspSerAsnIleThrGlu---------ThrIleGluAsnLeuArgAspGln 268

Qy        163 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
                                  |||          |||   :::|||
Db        269 LeuGluLysAlaThrAspGluGluHisLysLys-----------GluIleGluSerGln 284

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              ::::::   :::::::::   |||||:::|||   ::::::      :::|||::::::
Db        285 ValAspAlaLysLysLysGluLysGluGluLeuAspLysLysAlaIleAsnLeuAspLys 304

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
                 |||::::::||||||         :::   :::|||:::   :::|||
Db        305 ---------AlaGlnGlnLysLeuAspSerAlaGluAspAsnLeuAspValGlnArgAsp 321

Qy        340 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 399
                 :::||||||:::||||||      :::::::::   :::|||:::
Db        322 ThrValArgGluLysIleGlnGluAspIleAsnGluIleAsnLysGluLysAsnLeuPro 341

Qy        400 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 459
              :::                        ::::::::::::         ||||||   |||
Db        342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy        460 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 519
              ::::::   |||      |||::::::                  :::|||   :::|||
Db        362 GluAspLeuGlnGluGlnLeuLysGluAlaGlyAspGluAsnGlnLys---ArgGluIle 380

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              :::      :::|||   :::::::|||   :::|||   |||:::   :::
Db        381 GluLysGlnIleGluIleLysLysArgAspGluGluLeuLeuLysSerLysAspGlyLys 400

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              :::   ::::::      |||   :::                        |||   |||
Db        401 ValSerLysAspTyrGluAlaLeuAsp-----------------LeuAspArgGluLeu 414

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                 ||||||   :::|||:::      :::   |||:::|||
Db        415 SerLysAlaSerSerLysGluLysSerLysValLysGluGlu----------------- 428

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||         |||:::                  :::|||         |||::::::
Db        429 GluIleThrLysGlyLysSerArgAlaSerLeuGlyAspLeu------AsnAsnAspLys 446

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT------ 813
                 :::|||            ||||||:::|||   :::   :::|||
Db        447 AsnLeuMetLeuProGluAspGlnLysLeuProGluAspLysLysLeuAspSerLysLeu 466

Qy        814 ---------AGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATT 864
                    ::::::   |||   :::|||   :::||||||   ::::::   |||
Db        467 AspGlyLysLysGluPheLysProValSerGluValGluLysLeuAspLysIleSerLys 486

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
```

```
              :::   |||   |||
Db        487 SerAsnAsnAsnGlu-----------------------------------------Val 492

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAA--------------GATAAATCTGCTGACATA 969
              |||   :::   |||      |||                  |||         ::::::
Db        493 GlyLysLeuSerProLeuAspLysProSerTyrAspAspIleAspSerLysGluGluVal 512

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
                 |||        :::  :::|||:::     ::::::                |||
Db        513 AspAsn---------LysAlaIleAsnLeuGlnLysIleAsp------------ProLys 525

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                 :::|||:::   ::::::      :::|||      :::|||::::::
Db        526 ValLysAspGlnThrThrSerLeuAsnGluAsp------LeuAspLysAspLeuThrThr 543

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
                    :::|||         |||
Db        544 MetSerIleAspSerSerSerPro---------------------------------- 551

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT------ 1203
                 ::::::   |||:::   |||   |||                  |||
Db        552 ------ValPheLeuGluValIleAspProIleThrAsnLeuGlyThrLeuGlnLeuIle 569

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263
                 ||||||:::   |||:::            :::|||   :::::::::   |||::::::
Db        570 AspLeuAsnThrGlyValArg------LeuLysGluSerThrGlnGlnGlyIleGlnArg 587

Qy       1264 GGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG------------AT 1311
                    ||||||:::   ||||||  :::::::::   :::|||
Db        588 TyrGlyIleTyrGluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLysAl 607

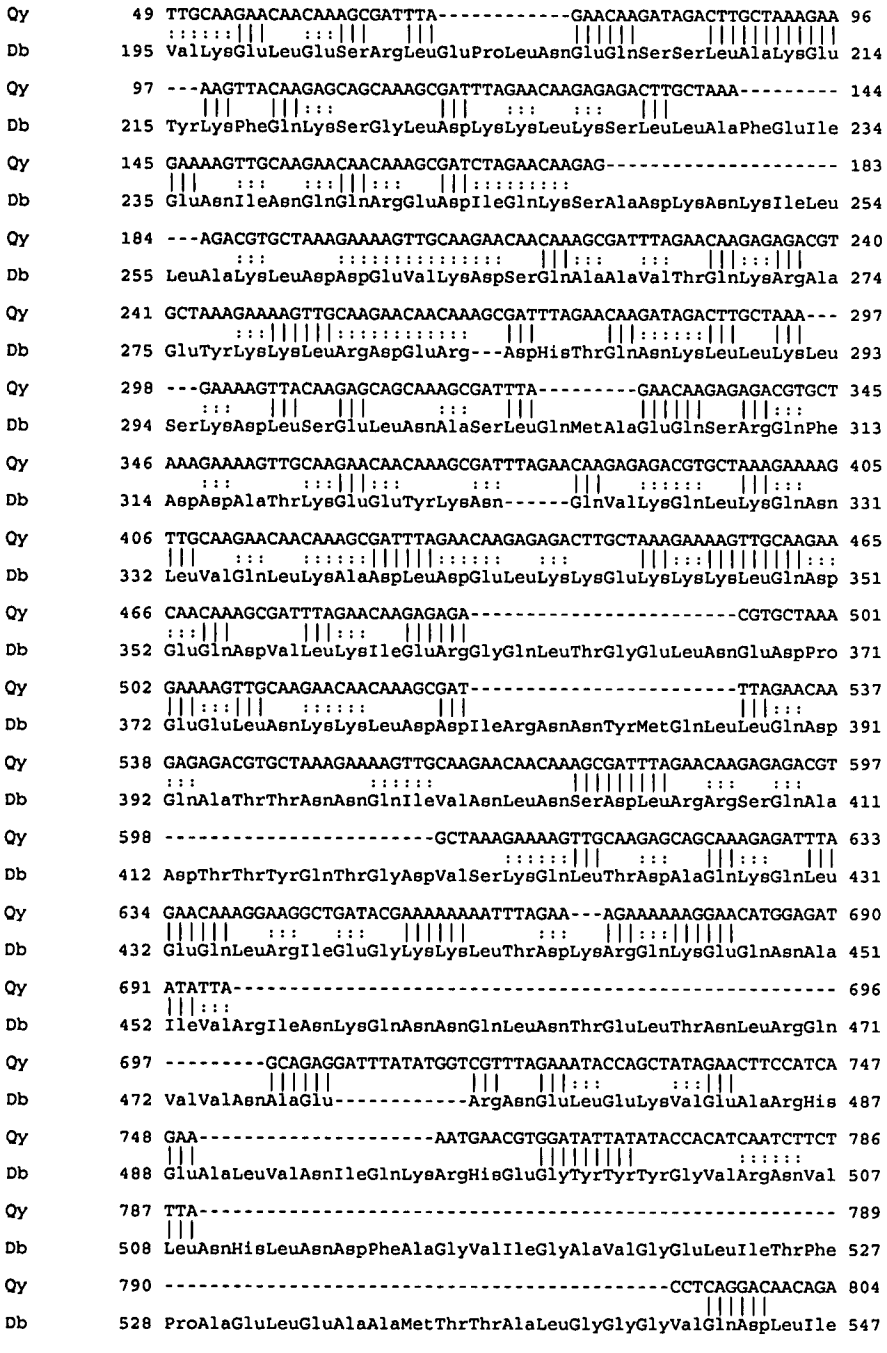
Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
                 |||:::   :::::::::   |||:::|||:::|||::::::|||
Db        607 aLysLeuGlnIleLeuAsnLysLeuGluAsnLeuLysValValSer 622

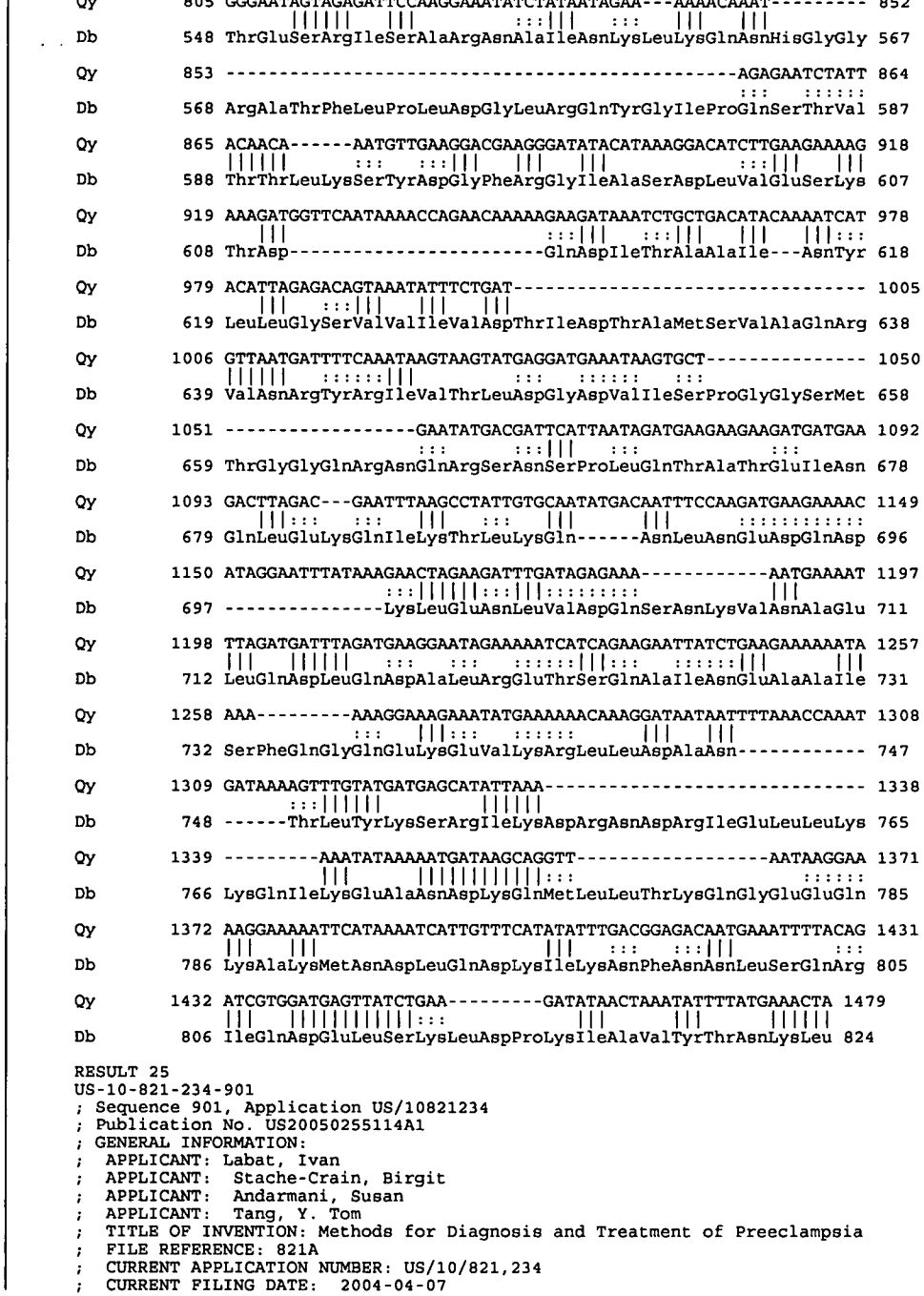
RESULT 24
US-11-074-176-134
; Sequence 134, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
;  APPLICANT: Klaenhammer, Todd R.
;  APPLICANT:  Russell, William M.
;  APPLICANT:  Altermann, Eric
;  APPLICANT:  McAuliffe, Olivia
;  APPLICANT:  Peril, Andrea Azcarate
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding
;  TITLE OF INVENTION:  Stress-Related Proteins and Uses Therefore
;  FILE REFERENCE: 5051-694
;  CURRENT APPLICATION NUMBER: US/11/074,176
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: 60/551,161
;  PRIOR FILING DATE: 2004-03-08
;  NUMBER OF SEQ ID NOS: 381
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1189
;   TYPE: PRT
;   ORGANISM: Lactobacillus acidophilus
US-11-074-176-134

Alignment Scores:
Pred. No.:              3.74e-06        Length:         1189
Score:                  230.50          Matches:        144
Percent Similarity:     40.21%          Conservative:   129
Best Local Similarity:  21.21%          Mismatches:     191
Query Match:            9.27%           Indels:         215
DB:                     7               Gaps:           33

US-09-837-344-41 (1-1482) x US-11-074-176-134 (1-1189)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGA-----------CGTGCTAAAGAAAAG 48
              :::|||   |||:::   |||::::::   :::            |||     :::
Db        175 GluGluAlaGlnAsnGlnLeuLysLysThrGlnAspAsnLeuIleArgIleAsnAspLeu 194
```





RESULT 25
US-10-821-234-901
; Sequence 901, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07

```
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 901
;   LENGTH: 1586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-901

Alignment Scores:
Pred. No.:                4.8e-06        Length:          1586
Score:                    228.00         Matches:         126
Percent Similarity:       38.30%         Conservative:    95
Best Local Similarity:    21.84%         Mismatches:      188
Query Match:              9.17%          Indels:          168
DB:                       6              Gaps:            23

US-09-837-344-41 (1-1482) x US-10-821-234-901 (1-1586)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAA 54
              ::::::|||:::   ||||||   :::    |||  ||||||    ||||||
Db       1025 GluGlnGlnArgLysAlaLeuGluAlaLysAlaAlaAlaPheGluLysGlnValLeuGln 1044

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG------ 108
                |||   |||      |||   :::        ||||||||||   |||
Db       1045 LeuGlnAlaSerHisArgGluSerGluGluAlaLeuGlnLysArgLeuAspGluValSer 1064

Qy        109 ---------------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 153
                           ||||||                 |||         ||||||
Db       1065 ArgGluLeuCysHisThrGlnSerSerHisAlaSerLeuArgAlaAspAlaGluLysAla 1084

Qy        154 CAAGAACAACAAAGCGATCTA----------------------------GAACAAGAG 183
              ||||||||||||      :::                             |||   |||
Db       1085 GlnGluGlnGlnGlnMetAlaGluLeuHisSerLysLeuGlnSerSerGluAlaGlu 1104

Qy        184 AGACGTGCTAAA---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||:::|||   |||:::|||                 |||::::::   |||
Db       1105 ValArgSerLysCysGluGluLeuSerGlyLeuHisGlyGlnLeuGlnGluAlaArgAla 1124

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA----------------CAA 282
                  :::|||   |||:::         :::|||                   |||
Db       1125 GluAsnSerGlnLeuThrGluArgIleArgSerIleGluAlaLeuLeuGluAlaGlyGln 1144

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
              |||   |||::::::   :::|||      |||::::::   :::|||::::::
Db       1145 AlaArgAspAlaGlnAsp---ValGlnAlaSerGlnAlaGluAlaAspGlnGlnGlnThr 1163

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-----------AGA 390
                :::::|||:::   |||   |||   ||||||:::|||                |||
Db       1164 ArgLeuLysGluLeuGluSerGlnValSerGlyLeuGluLysGluAlaIleGluLeuArg 1183

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA-------------- 435
              |||   |||:::   :::   :::   :::||||||   :::
Db       1184 GluAlaValGluGlnGlnLysValLysAsnAsnAspLeuArgGluLysAsnTrpLysAla 1203

Qy        436 --------------------GAGAGACTTGCTAAAGAAAAGTTGCAAGAA--------- 465
                                 |||:::     ||||||||||
Db       1204 MetGluAlaLeuAlaThrAlaGluGlnAlaCysLysGluLysLeuHisSerLeuThrGln 1223

Qy        465 ------------------------------------------------------------ 465

Db       1224 AlaLysGluGluSerGluLysGlnLeuCysLeuIleGluAlaGlnThrMetGluAlaLeu 1243

Qy        466 --------------------------------CAACAAAGCGATTTAGAACAAGAGAGA 492
                                              ||||||:::      |||      :::
Db       1244 LeuAlaLeuLeuProGluLeuSerValLeuAlaGlnGlnAsnTyrThrGluTrpLeuGln 1263

Qy        493 CGTGCTAAAGAAAAG----------------------------TTGCAAGAACAACAA 522
                 |||||||||                              |||
Db       1264 AspLeuLysGluLysGlyProThrLeuLeuLysHisProProAlaProAlaGluProSer 1283

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---CAAAGCGAT 579
                 |||||||||      :::   |||      |||:::   |||         |||::::::
Db       1284 SerAspLeuAlaSerLysLeuArgGluAlaGluGluThrGlnSerThrLeuGlnAlaGlu 1303

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                 :::|||   |||            :::   :::          |||||||||   |||
Db       1304 CysAspGlnTyrArgSerIleLeuAlaGluThrGluGlyMetLeuArgAspLeu---Gln 1322

Qy        640 AGGAAGGCTGATACGAAAAAAAAATTTAGAAAGAAAAAAG---------------------- 678
                 :::      :::   ::::::   :::   |||   |||
Db       1323 LysSerValGluGluGluGluGlnValTrpArgAlaLysValGlyAlaAlaGluGluGlu 1342

Qy        679 ----------------------GAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                                     :::|||   :::   :::   |||   |||   |||
Db       1343 LeuGlnLysSerArgValThrValLysHisLeuGluGluIleValGluLysLeuLysGly 1362

Qy        715 CGTTTAGAAATACCA-----------------GCTATAGAACTTCCATCAGAAAATGAA 756
                 ||||||                      :::   |||   :::|||   |||
Db       1363 GluLeuGluSerSerAspGlnValArgGluHisThrLeuHisLeuGluAlaGluLeuGlu 1382

Qy        756 ------------------------------------------------------------ 756

Db       1383 LysHisMetAlaAlaAlaSerAlaGluCysGlnAsnTyrAlaLysGluValAlaGlyLeu 1402

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
                 |||        :::      ||||||   |||                   :::
Db       1403 ArgGlnLeuLeuLeuGluSerGlnSerGlnLeuAspAlaAlaLysSerGluAlaGlnLys 1422

Qy        817 GATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT 876
                 |||   |||:::::::::::   :::          |||
Db       1423 GlnSerAspGluLeuAlaLeuValArgGlnGlnLeuSerGluMet--------------- 1437

Qy        877 GAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA--- 933
                                  |||   |||:::|||          ||||||   |||
Db       1438 --------------------LysSerHisValGlu---------AspGlyAspIleAla 1447

Qy        934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                              ||||||          :::        :::|||
Db       1448 GlyAlaProAlaSerSerProGluAlaProProAlaGluGlnAspProValGlnLeuLys 1467

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
                 |||                 :::   :::   |||   |||      :::
Db       1468 ThrGlnLeuGluTrpThrGluAlaIleLeuGluAspGluGlnThrGlnArg------Gln 1485

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTA 1098
              :::::::::||||||::::::::::::                          |||:::|||
Db       1486 LysLeuThrAlaGluPheGluGluAlaGlnThrSerAlaCysArgLeuGlnGluGluLeu 1505

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
              ::::::   :::                 ::::::   :::   ||||||
Db       1506 GluLysLeuArgThrAlaGlyProLeuGluSerSerGluThrGluGlu------------ 1521

Qy       1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------ 1206
                   :::|||::::::   :::||||||   :::   |||
Db       1522 AlaSerGlnLeuLysGluArgLeuGluLysGluLysLysLeuThrSerAspLeuGlyArg 1541

Qy       1207 ------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                         |||   |||   ::::::   :::   |||:::|||:::   ||||||
Db       1542 AlaAlaThrArgLeuGlnGluLeuLeuLysThrThrGlnGluGlnLeuAlaArgGluLys 1561

Qy       1255 ------ATAAAAAAA---GGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                    :::||||||      ::::::   ||||||   :::|||   :::
Db       1562 AspThrValLysLysLeuGlnGluGlnLeuGluLysAlaGluAspGlySer 1578

RESULT 26
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
```

```
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-68

Alignment Scores:
Pred. No.:              6.15e-06        Length:         693
Score:                  227.00          Matches:        98
Percent Similarity:     43.82%          Conservative:   111
Best Local Similarity:  20.55%          Mismatches:     143
Query Match:            9.13%           Indels:         125
DB:                     7               Gaps:           19

US-09-837-344-41 (1-1482) x US-11-196-475-68 (1-693)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
Db        230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy         73 GAACAAGATAGACTTGCTAAA----------------GAAAAGTTACAAGAGCAGCAA 114
Db        247 GluLeuAspAsnValHisLysSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        267 AspGlnLeuGlu---LysAlaThrAspGluGluHisArgLysGluIleGluSerGlnVal 285

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
Db        286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---CTT 291
Db        304 ---LysAlaGlnGlnLysLeuAspSerSerGluAspAsnLeuAspIleGlnArgAspThr 322

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
Db        323 ValArgGluLysIleGlnGluAspIleAspGluIleAsnLysGluLysAsnLeuProLys 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
Db        343 ProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeuGlu 362

Qy        412 GAACAACAA----------------AGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
Db        363 AspLeuGlnGluGlnLeuLysGluThrSerAspGluAsnGlnLysArgGluIleGluLys 382

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
Db        383 GlnIleGluIleLysLysSerAspGluGluLeuLeuLysSerLysAspProLysAlaLeu 402

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 573
Db        403 AspLeuAsnGlyAspLeuAsnSerLysValSerSerLysGluLysIleLysGlyLysGlu 422

Qy        574 AGCGATTTA---GAACAAGAGAGACGTGCTAAA----------GAAAAG 609
Db        423 GlyGluIleValLysGluGluSerLysAlaSerLeuAlaAspLeuAsnAsnAspGluAsn 442

Qy        610 TTGCAA------GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAT 663
Db        443 LeuMetArgProGluAspGlnLysLeuSerGluAspLysLysLeuAspSerLysLysAsn 462

Qy        664 TTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA 723
Db        463 LeuLys---------------------------------------------------- 464

Qy        724 ATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCT 783
Db        465 ---ProValSerGluIleGluArgValAsnGlu-------------------------- 474

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA 843
Db        475 ------------IleSerLysSerAsnAsnAsnGluIleSerGluSerSerProLeuTyr 490

Qy        844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903
Db        491 LysProSer------------TyrSerAspMetAspSerLysGluGlyIle--------- 503

Qy        904 CATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
Db        504 ---------AspAsnLysAspValAsnLeuGlnGluThrLysSerGlnThrLysSerGln 520

Qy        964 GACATACAA---AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        521 ProThrSerLeuAsnGlnAspLeuThrThrMetSerIleAspSerSerAsnProVal--- 539

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        539 ---------------------------------------------------------- 539

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db        540 ------------PheLeuGluValIleAspProIleThr---------------------- 548

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db        549 ------AsnLeuGlyThrLeuGln--------------------------------Leu 555

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
Db        556 IleAspLeuAsnThrGlyValArg------LeuLysGluSerThrGlnGlnGlyIleGln 573

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG----------- 1309
Db        574 ArgTyrGlyIleTyrGluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLy 593

Qy       1310 -ATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
Db        593 sAlaLysLeuGlnIleLeuAsnLysLeuGluAsnLeuLysValIleSer 609

RESULT 27
US-10-878-556A-136
; Sequence 136, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 136
;   LENGTH: 1410
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_hum/rrb1_human
```

```
;   DATABASE ENTRY DATE: 2003-02-28
US-10-878-556A-136

Alignment Scores:
Pred. No.:              8.81e-06        Length:         1410
Score:                  223.00          Matches:        122
Percent Similarity:     39.13%          Conservative:   103
Best Local Similarity:  21.22%          Mismatches:     186
Query Match:            8.97%           Indels:         164
DB:                     6               Gaps:           23

US-09-837-344-41 (1-1482) x US-10-878-556A-136 (1-1410)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAA 54
              ::::::|||:::      ||||||   :::      |||   ||||||      ||||||
Db        849 GluGlnGlnArgLysAlaLeuGluAlaLysAlaAlaAlaPheGluLysGlnValLeuGln 868

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG------ 108
                |||   |||      |||   :::          ::::::::::|||   |||
Db        869 LeuGlnAlaSerHisArgGluSerGluGluAlaLeuGlnLysArgLeuAspGluValSer 888

Qy        109 ---------------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 153
                           ||||||              |||         ||||||
Db        889 ArgGluLeuCysHisThrGlnSerSerHisAlaSerLeuArgAlaAspAlaGluLysAla 908

Qy        154 CAAGAACAACAAAGCGATCTA----------------------------GAACAAGAG 183
              ||||||||||||      :::                             |||   |||
Db        909 GlnGluGlnGlnGlnMetAlaGluLeuHisSerLysLeuGlnSerSerGluAlaGlu 928

Qy        184 AGACGTGCTAAA-----------------------GAAAAGTTGCAAGAACAACAAAGC 219
              |||:::|||                       :::|||||||||   ::::::
Db        929 ValArgSerLysCysGluGluLeuSerGlyLeuHisGlyGlnLeuGlnGluAlaArgAla 948

Qy        220 GATTTAGAACAA-----------------------------------------------GAG 234
              :::      |||                                              :::
Db        949 GluAsnSerGlnLeuThrGluArgIleArgSerIleGluAlaLeuLeuGluAlaGlyGln 968

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 294
                |||      :::   :::|||      |||::::::   :::|||   :::
Db        969 AlaArgAspAlaGlnAspValGlnAlaSerGlnAlaGluAlaAspGlnGlnGlnThrArg 988

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG------------AGACGT 342
              ::::::|||:::   |||   |||   ||||||:::|||            |||
Db        989 LeuLysGluLeuGluSerGlnValSerGlyLeuGluLysGluAlaIleGluLeuArgGlu 1008

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA------------------ 384
              |||   |||:::   :::   :::   :::||||||   :::
Db       1009 AlaValGluGlnGlnLysValLysAsnAsnAspLeuArgGluLysAsnTrpLysAlaMet 1028

Qy        385 ------------------GAGAGACGTGCTAAAGAAAAGTTG------------------ 408
                              |||:::      ||||||||||||
Db       1029 GluAlaLeuAlaThrAlaGluGlnAlaCysLysGluLysLeuLeuSerLeuThrGlnAla 1048

Qy        409 ---CAAGAACAACAAAGCGATTTA---------------------------------- 429
                  :::|||   :::      |||
Db       1049 LysGluGluSerGluLysGlnLeuCysLeuIleGluAlaGlnThrMetGluAlaLeuLeu 1068

Qy        430 ------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
                        |||         ||||||:::::          |||          |||
Db       1069 AlaLeuLeuProGluLeuSerValLeuAlaGlnGlnAsnTyrThrGluTrpLeuGlnAsp 1088

Qy        478 TTAGAACAAGAG------------------------------------------------ 489
              |||:::::::::
Db       1089 LeuLysGluLysGlyProThrLeuLeuLysHisProProAlaProAlaGluProSerSer 1108

Qy        490 ------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531
                            |||   |||:::|||              |||::::::
Db       1109 AspLeuAlaSerLysLeuArgGluAlaGluGluThrGlnSerThrLeuGlnAlaGluCys 1128

Qy        532 GAACAAGAGAGA------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              :::|||   |||               :::      |||::::::   |||
Db       1129 AspGlnTyrArgSerIleLeuAlaGluThrGluGlyMetLeuArgAspLeuGlnLysSer 1148

Qy        580 TTAGAACAAGAGAGACGT----------------------GCTAAAGAAAAGTTGCAA 615
              :::|||:::|||   :::                      |||:::|||:::||||||
Db       1149 ValGluGluGluGluGlnValTrpArgAlaLysValGlyAlaAlaGluGluGluLeuGln 1168

Qy        616 GAGCAGCAA---------AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::         :::   ||||||:::                 |||      |||
Db       1169 LysSerArgValThrValLysHisLeuGluGluIleValGluLysLeuLysGlyGluLeu 1188

Qy        667 GAA---------AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              |||         :::   :::||||||          |||      :::|||
Db       1189 GluSerSerAspGlnValArgGluHisThrSerHisLeuGluAlaGluLeuGluLysHis 1208

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA--------------CGTGGA 762
              :::         :::   |||      :::      :::                |||
Db       1209 MetAlaAlaAlaSerAlaGluCysGlnAsnTyrAlaLysGluValAlaGlyLeuArgGln 1228

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC 822
                  :::      ||||||   |||                           :::   |||
Db       1229 LeuLeuLeuGluSerGlnSerGlnLeuAspAlaAlaLysSerGluAlaGlnLysGlnSer 1248

Qy        823 AAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGA 882
              |||:::::::::::::   :::        |||
Db       1249 AspGluLeuAlaLeuValArgGlnGlnLeuSerGluMet---------------------- 1261

Qy        883 CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA--------- 933
                          |||   |||:::|||          ||||||   |||
Db       1262 --------------LysSerHisValGlu---------AspGlyAspIleAlaGlyAla 1273

Qy        934 ---------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                        ||||||           :::         :::|||      |||
Db       1274 ProAlaSerSerProGluAlaProProAlaGluGlnAspProValGlnLeuLysThrGln 1293

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                          :::   :::   |||   |||      :::          ::::::
Db       1294 LeuGluTrpThrGluAlaIleLeuGluAspGluGlnThrGlnArg------GlnLysLeu 1311

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
              :::||||||:::::::::::                        |||:::|||::::::
Db       1312 ThrAlaGluPheGluGluAlaGlnThrSerAlaCysArgLeuGlnGluGluLeuGluLys 1331

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                 :::                 ::::::   :::   ||||||
Db       1332 LeuArgThrAlaGlyProLeuGluSerSerGluThrGluGlu------------AlaSer 1347

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------------ 1206
              :::|||::::::   :::||||||   :::   |||
Db       1348 GlnLeuLysGluArgLeuGluLysGluLysLysLeuThrSerAspLeuGlyArgAlaAla 1367

Qy       1207 ------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA------ 1254
                    |||   |||   ::::::   :::   |||:::|||:::   ||||||
Db       1368 ThrArgLeuGlnGluLeuLeuLysThrThrGlnGluGlnLeuAlaArgGluLysAspThr 1387

Qy       1255 ATAAAAAAA---GGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
              :::||||||      ::::::   ||||||   :::|||   :::
Db       1388 ValLysLysLeuGlnGluGlnLeuGluLysAlaGluAspGlySer 1402

RESULT 28
US-10-821-234-975
; Sequence 975, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
```

```
; SEQ ID NO 975
;   LENGTH: 989
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-975

Alignment Scores:
Pred. No.:                1.25e-05        Length:          989
Score:                    220.50          Matches:         116
Percent Similarity:       41.77%          Conservative:    115
Best Local Similarity:    20.98%          Mismatches:      210
Query Match:              8.87%           Indels:          112
DB:                       6               Gaps:            21

US-09-837-344-41 (1-1482) x US-10-821-234-975 (1-989)

Qy         31 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 90
              ::::::   :::|||   ||||||:::  |||      ||||||
Db        441 GlnGlnLeuGlnGluIleLeuGlnAspLeuGlnLysArgLeuGluSerSerGluAlaGlu 460

Qy         91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA------CTTGCTAAA 144
              :::::::::||||||   :::                  |||         |||
Db        461 ArgLysGlnLeuGlnValGluLeuGlnSerArgArgAlaGluLeuValCysLeuAsnAsn 480

Qy        145 GAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 204
              ::::::   |||      |||||||||   |||                  ||||||
Db        481 ThrGluIleSerGluAsnSerSerAspLeuSerGln----------------LysLeu 494

Qy        205 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 264
              :::|||   ||||||      |||:::   :::          :::|||:::|||
Db        495 LysGluThrGlnSerLysTyrGluGluAlaMetLysGluValLeuSerValGlnLysGln 514

Qy        265 CAA-----------------------------------AGCGATTTAGAACAAGAT 285
                                                    |||         :::
Db        515 MetLysLeuGlyLeuValSerProGluSerMetAspAsnTyrSerHisPheHisGluLeu 534

Qy        286 AGACTTGCTAAAGAAAAGTTA----------CAAGAGCAGCAAAGCGATTTAGAACAA 333
              |||:::   :::|||::::::          |||:::   |||:::   ||||||:::
Db        535 ArgValThrGluGluGluIleAsnValLeuLysGlnAspLeuGlnAsnAlaLeuGluGlu 554

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 393
                 |||   |||||||||::::::|||   :::      |||   :::      :::
Db        555 SerGluArgAsnLysGluLysValArgGluLeuGluGluLysLeuValGluArgGluLys 574

Qy        394 GCTAAA----------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
                                    |||:::   :::|||   :::|||
Db        575 GlyThrValIleLysProProValGluGluTyrGluGluMetLysSerSerTyrCysSer 594

Qy        436 ------GAGAGACTTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGATTTA 480
                    |||   :::   |||||||||          ::::::   |||
Db        595 ValIleGluAsnMetAsnLysGluLysAlaPheLeuPheGluLysTyrGlnGluAlaGln 614

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::      :::   |||:::   |||:::   |||   :::
Db        615 GluGluIleMetLysLeuLysAspThrLeuLysSerGlnMetThr--------------- 629

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                       ::|||          |||   :::          |||          |||
Db        630 ---------GlnGluAlaSerAspGluAlaGluAspMetLysGluAlaMetAsnArgMet 646

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
                 ::::::|||   :::|||      :::|||   |||          :::   :::
Db        647 IleAspGluLeuAsnLysGlnValSerGluLeuSerGlnLeuTyrLysGluAlaGlnAla 666

Qy        661 AATTTAGAA---------AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711
              ||||||           :::::::|||          |||:::   ||||||   ::::::
Db        667 GluLeuGluAspTyrArgLysArgLysSerLeuGluAspValThrAlaGluTyrIleHis 686

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771
                                          :::|||:::|||:::         :::
Db        687 ----------------------------LysAlaGluHisGluLysLeuMetGlnLeu 696

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC--------- 822
```

```
                 :::   |||           ::::::      :::         |||
Db        697 ThrAsnValSerArgAlaLysAlaGluAspAlaLeuSerGluMetLysSerGlnTyrSer 716

Qy        823 --------------------------AAGGAAATATCTATAATAGAAAAAACAAATAGA 855
                                         |||::::::          :::|||   |||
Db        717 LysValLeuAsnGluLeuThrGlnLeuLysGlnLeuValAspAlaGlnLysGluAsnSer 736

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA--- 912
              |||||||||   :::::::::          :::                  |||
Db        737 ValSerIleThrGluHisLeuGlnValIleThrThrLeuArgThrAlaAlaLysGluMet 756

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAA------CAAAAAGAAGATAAATCTGCTGAC 966
              |||:::|||      :::::::|||                |||||   :::   |||
Db        757 GluGluLysIleSerAsnLeuLysGluHisLeuAlaSerLysGluValGluValAlaLys 776

Qy        967 ATACAAAATCATACATTAGAG----------------------ACAGTAAATATTTCT 1002
              ::::::          ||||||                        |||     |||
Db        777 LeuGluLysGlnLeuLeuGluGluLysAlaAlaMetThrAspAlaMetValProArgSer 796

Qy       1003 GATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT---------GCTGAA 1053
                          |||   |||      |||   |||:::|||         ::::::
Db        797 SerTyrGluLysLeuGln---SerSerLeuGluSerGluValSerValLeuAlaSerLys 815

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
                    :::|||:::   ::::::|||:::         ::::::   :::   :::
Db        816 LeuLysGluSerValLysGluLysGluLysValHisSerGluValValGlnIleArgSer 835

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173
                 |||                 :::::::::|||||||||
Db        836 GluValSerGlnValLys---ArgGluLysGluAsnIleGlnThrLeu------------ 850

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
                 |||      |||   :::   :::::::::|||   :::      :::::::::
Db        851 ---LeuLysSerLysGluGlnGluValAsnGluLeuLeuGlnLysPheGlnGlnAlaGln 869

Qy       1234 GAAGAATTATCTGAAGAAAAA------ATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
              |||||||||:::|||   |||         :::         |||   |||:::   |||
Db        870 GluGluLeuAlaGluMetLysArgTyrAlaGluSerSerSerLysLeuGluGluAspLys 889

Qy       1288 GATAATAATTTTAAACCAAATG-----------------ATAAAAGTTTGTATGATGAG 1329
              |||      :::|||:::|||                 :::|||      :::   ||
Db        890 AspLysLys-IleAsnGluMetSerLysGluValThrLysLeuLysGluAlaLeuAsnSe 909

Qy       1330 CATA-----------------------------------TTAAAAAATATAAAAAT 1350
              |:::                                 :::   ::::::
Db        909 rLeuSerGlnLeuSerTyrSerThrSerSerSerLysArgGlnSerGlnGlnLeuGluAl 929

Qy       1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGAC 1410
                :::   :::   :::::::::   :::|||            |||
Db        929 aLeuGlnGlnGlnValLysGlnLeuGlnAsnGlnLeuAlaGluCysLysLysGlnHisGl 949

Qy       1411 GGAGACAATGAAATTTTACAGATCGTGGATGAGTTAT 1447
                |||   :::      ||||||      :::   |||
Db        949 nGluValIleSerValTyrArgMetHisLeuLeuTyr 961

RESULT 29
US-11-196-475-74
; Sequence 74, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
```

```
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Alignment Scores:
Pred. No.:                1.48e-05        Length:          700
Score:                    219.50          Matches:         102
Percent Similarity:       46.19%          Conservative:    116
Best Local Similarity:    21.61%          Mismatches:      144
Query Match:              8.83%           Indels:          110
DB:                       7               Gaps:            25

US-09-837-344-41 (1-1482) x US-11-196-475-74 (1-700)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
Db        230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy         73 GAACAAGATAGACTTGCTAAA-----------------GAAAAGTTACAAGAGCAGCAA 114
Db        247 GluLeuAspAsnIleHisGluSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        267 AspGlnLeuGlu---LysAlaThrAspGluGluHisLysLysGluIleGluSerGlnVal 285

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
Db        286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy        235 AGACGTGCTAAAGAAAAGTTGCAA------GAACAACAAAGCGATTTAGAACAAGATAGA 288
Db        304 ---LysAlaGlnGlnLysLeuAspPheAlaGluAspAsnLeuAspIleGlnArgAspThr 322

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
Db        323 Val---ArgGluLysLeuGlnGluAsnIleAsnGluThrAsnLysGluLysAsnLeuPro 341

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
Db        342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
Db        362 GluAspLeuGlnGluGlnLeu-------------------------------LysGlu 370

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
Db        371 ThrGlyAspGluAsnGlnLysArgGluIleGluLysGlnIleGluIleLysLysSerAsp 390

Qy        529 ------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
Db        391 GluLysLeuLeuLysSerLysAspAspLysAlaSerLysAspGlyLysAlaLeuAspLeu 410

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
Db        411 AspArgGluLeuAsnSerLysAlaSerSerLysGluLysSerLysAlaLysGluGluGlu 430

Qy        643 ------AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696
Db        431 IleThrLysGlyLysSerGlnLysSerLeu-----------------GlyAspLeu--- 443

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756

Db        444 ----------------------------------------------------AsnAsnAsp 446

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT--- 813
Db        447 GluAsnLeuMetMetProGluAspGlnLysLeuProGluValLysLysLeuAspSerLys 466

Qy        814 AGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
Db        467 LysGluPheLysProValSerGluValGluLysLeuAspLys------------------ 480

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db        481 ------------------IlePheLysSerAsn---------AsnAsnValGlyGluLeu 491

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCT------GACATACAAAATCATACATTAGAG 987
Db        492 SerPro---------LeuAspLysSerSerTyrLysAspIleAspSerLysGlu---Glu 507

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAGGATGAA 1041
Db        508 ThrValAsn---LysAspValAsn---LeuGlnLysThrLysProGlnValLysAspGln 525

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
Db        526 ValThrSer-----------------LeuAsnGluAspLeuThrThrMetSerIleAsp 539

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        540 SerSerSerPro-----------------------------------------ValPhe 545

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT------GATTTAGATGAA 1215
Db        546 LeuGluValIleAspProIleThrAsnLeuGlyThrLeuGlnLeuIleAspLeuAsnThr 565

Qy       1216 GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
Db        566 GlyValArg------LeuLysGluSerThrGlnGlnGlyIleGlnArgTyrGlyIleTyr 583

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATG-----------ATAAAAGTTTGTAT 1323
Db        584 GluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLysAlaLysLeuGlnIl 603

Qy       1324 GATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
Db        603 eLeuAspLysLeuGluAsnLeuLysValValSer 614

RESULT 30
US-11-196-475-66
; Sequence 66, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Alignment Scores:

Pred. No.:	2.11e-05	Length:	700
Score:	216.50	Matches:	99
Percent Similarity:	46.07%	Conservative:	118
Best Local Similarity:	21.02%	Mismatches:	146
Query Match:	8.71%	Indels:	108
DB:	7	Gaps:	24

US-09-837-344-41 (1-1482) x US-11-196-475-66 (1-700)

```
Qy        13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
             :::|||::::::   |||          |||      :::   :::   |||:::
Db       230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy        73 GAACAAGATAGACTTGCTAAA----------------GAAAAGTTACAAGAGCAGCAA 114
             |||   |||   :::   :::              |||   ::::::      :::
Db       247 GluLeuAspAsnIleHisGluSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy       115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
             ||||||   :::         :::|||      :::|||   :::|||   :::
Db       267 AspGlnLeuGlu---LysAlaThrAspGluGluHisLysLysGluIleGluSerGlnVal 285

Qy       175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
             :::   :::::::::   ||||||:::|||   ::::::      :::|||:::
Db       286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAsnLeuAsp------ 303

Qy       235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGATAGACTT 291
                :::|||::::::||||||          :::   :::|||:::   |||
Db       304 ---LysAlaGlnGlnLysLeuAspSerAlaGluAspAsnLeuAspValGlnArgAsnThr 322

Qy       292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
             :::||||||:::||||||      :::::::::   :::|||:::         :::
Db       323 ValArgGluLysIleGlnGluAspIleAsnGluIleAsnLysGluLysAsnLeuProLys 342

Qy       352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
                                 ::::::::::::   :::   ||||||   |||:::
Db       343 ProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeuGlu 362

Qy       412 GAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 471
             :::   |||     |||                                  ::::::
Db       363 AspLeuGlnGluGlnLeu-------------------------------LysGluThr 371

Qy       472 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--- 528
                |||      |||:::|||      ::::::::::::::   ::::::||||||
Db       372 GlyAspGluAsnGlnLysArgGluIleGluLysGlnIleGluIleLysLysSerAspGlu 391

Qy       529 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                |||   :::   :::     |||          :::      ||||||:::
Db       392 LysLeuLeuLysSerLysAspAspLysAlaSerLysAspGlyLysAlaLeuAspLeuAsp 411

Qy       586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG--- 642
             :::|||      :::|||         :::|||:::   :::      |||:::
Db       412 ArgGluLeuAsnSerLysAlaSerSerLysGluLysSerLysAlaLysGluGluGluIle 431

Qy       643 ---AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                |||      :::::::|||:::|||                |||||:::
Db       432 ThrLysGlyLysSerGlnLysSerLeu-----------------GlyAspLeu------ 443

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
                                                          |||:::
Db       444 ------------------------------------------------AsnAsnAspGlu 447

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---AGA 816
                 :::|||            ||||||:::      :::   :::|||   :::
Db       448 AsnLeuMetMetProGluAspGlnLysLeuProGluValLysLysLeuAspSerLysLys 467

Qy       817 GATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT 876
             :::   |||   :::|||   :::||||||   ::::::
Db       468 GluPheLysProValSerGluValGluLysLeuAspLys---------------------- 480

Qy       877 GAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA 936
                           |||   |||   :::                  |||   :::
Db       481 ---------------IlePheLysSerAsn---------AsnAsnValGlyGluLeuSer 492

Qy       937 CCAGAACAAAAAGAAGATAAATCTGCT------GACATACAAAATCATACATTAGAGACA 990
             |||           |||||||||:::      ||||||   :::         ||||||
Db       493 Pro---------LeuAspLysSerSerTyrLysAspIleAspSerLysGlu---GluThr 508

Qy       991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAGGATGAAATA 1044
             ||||||      |||||||||      |||   :::|||         :::|||::::::
Db       509 ValAsn---LysAspValAsn---LeuGlnLysThrLysProGlnValLysAspGlnVal 526

Qy      1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
             ::::::                 ::::::|||:::                ::::|||
Db       527 ThrSer-----------------LeuAsnGluAspLeuThrThrMetSerIleAspSer 540

Qy      1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                |||                                                ::::::
Db       541 SerSerPro-----------------------------------------ValPheLeu 546

Qy      1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT------GATTTAGATGAAGGA 1218
             |||:::   |||   |||                  |||         ||||||:::   |||
Db       547 GluValIleAspProIleThrAsnLeuGlyThrLeuGlnLeuIleAspLeuAsnThrGly 566

Qy      1219 ATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAA 1278
             :::              :::|||   :::::::::   |||::::::         ||||||
Db       567 ValArg------LeuLysGluSerThrGlnGlnGlyIleGlnArgTyrGlyIleTyrGlu 584

Qy      1279 AAAACAAAGGATAATAATTTTAAACCAAATG-----------ATAAAAGTTTGTATGAT 1326
             :::   ||||||  :::::::::   :::|||               |||:::   :::::
Db       585 ArgGluLysAsp-LeuValValIleLysMetAspSerGlyLysAlaLysLeuGlnIleLe 604

Qy      1327 GAGCATATTAAAAAATATAAAAATGATAAGC 1357
             :      |||:::|||:::|||::::::|||
Db       604 uAspLysLeuGluAsnLeuLysValValSer 614
```

Search completed: December 21, 2005, 11:01:54
Job time : 103.432 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:27:59 ; Search time 8587.96 Seconds
(without alignments)
9809.312 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1482	100.0	1482	6	AR162970	AR162970 Sequence
2	1482	100.0	1482	6	AR178259	AR178259 Sequence
3	1482	100.0	1496	2	PFLSA13	Z30320 P.falciparu
4	1480.4	99.9	1482	6	AR162974	AR162974 Sequence
5	1480.4	99.9	1482	6	AR178263	AR178263 Sequence
6	1446	97.6	1493	6	A28743	A28743 cDNA for LS
7	1446	97.6	1493	6	AR162967	AR162967 Sequence
8	1446	97.6	1493	6	AR178256	AR178256 Sequence
9	1442	97.3	5970	2	PFLSA1G	X56203 P.falciparu
10	1406.8	94.9	253001	2	AE014834	AE014834 Plasmodiu
11	950	64.1	950	6	A28740	A28740 cDNA for LS
12	950	64.1	950	6	AR162961	AR162961 Sequence
13	950	64.1	950	6	AR178250	AR178250 Sequence
14	843	56.9	843	2	PFALSAC	L40885 Plasmodium
15	841.4	56.8	843	2	PFALSAB	L40884 Plasmodium
16	841.4	56.8	843	2	PFALSAG	L40889 Plasmodium
17	841.4	56.8	843	2	PFALSAH	L40890 Plasmodium
18	841.4	56.8	843	2	PFALSAI	L40891 Plasmodium
19	839.8	56.7	843	2	PFALSAE	L40887 Plasmodium
20	839.8	56.7	843	2	PFALSAF	L40888 Plasmodium
21	839.8	56.7	843	2	PFALSAK	L40893 Plasmodium
22	839.8	56.7	843	2	PFALSAL	L40834 Plasmodium
23	838.2	56.6	843	2	PFALSAM	L40835 Plasmodium
24	836.6	56.5	843	2	PFALSA1N	L40947 Plasmodium
25	836.6	56.5	843	2	PFALSAD	L40886 Plasmodium
26	836.6	56.5	843	2	PFALSAN	L40836 Plasmodium
27	836.6	56.5	843	2	PFALSAO	L40837 Plasmodium
28	835	56.3	843	2	PFALSA1C	L40910 Plasmodium
29	835	56.3	843	2	PFALSAJ	L40892 Plasmodium
30	833.4	56.2	843	2	PFALSA1A	L40908 Plasmodium
31	833.4	56.2	843	2	PFALSA1B	L40909 Plasmodium
32	791.8	53.4	795	2	AF246996	AF246996 Plasmodiu
33	503.4	34.0	1368	11	AY751501	AY751501 Synthetic
34	472.8	31.9	988	2	PFLSA15	Z30319 P.falciparu
35	472.8	31.9	988	6	A28742	A28742 cDNA for LS
36	472.8	31.9	988	6	AR162963	AR162963 Sequence
37	472.8	31.9	988	6	AR178252	AR178252 Sequence
38	470.8	31.8	954	6	AR162966	AR162966 Sequence
39	470.8	31.8	954	6	AR178255	AR178255 Sequence
c 40	280.6	18.9	198224	9	AC149597	AC149597 Mus muscu
41	274.4	18.5	208332	9	AC122227	AC122227 Mus muscu
c 42	252.4	17.0	172716	14	AC115383	AC115383 Rattus no
43	250.8	16.9	184015	9	AC144849	AC144849 Mus muscu
c 44	237.4	16.0	161482	14	AC165336	AC165336 Mus muscu
45	225.8	15.2	2069	6	E10125	E10125 DNA encodin

ALIGNMENTS

RESULT 1
AR162970
LOCUS AR162970 1482 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 41 from patent US 6270771.
ACCESSION AR162970
VERSION AR162970.1 GI:16233435
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1482)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6270771-A 41 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1482
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1482; DB 6; Length 1482;
Best Local Similarity 100.0%; Pred. No. 1.3e-158;
Matches 1482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
```

```
Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
            |||||||||||||||||||||||||||||||||||||||||||
Db      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

RESULT 2
AR178259

LOCUS	AR178259	1482 bp	DNA	linear	PAT 20-APR-2002

DEFINITION Sequence 41 from patent US 6319502.
ACCESSION AR178259
VERSION AR178259.1 GI:20219397
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1482)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6319502-A 41 20-NOV-2001;
FEATURES Location/Qualifiers
source 1. .1482
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1482; DB 6; Length 1482;
Best Local Similarity 100.0%; Pred. No. 1.3e-158;
Matches 1482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy        61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
```

```
Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

```
RESULT 3
PFLSA13
LOCUS       PFLSA13                 1496 bp    DNA     linear   INV 05-SEP-1994
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen-1 (3').
ACCESSION   Z30320
VERSION     Z30320.1  GI:510183
KEYWORDS    liver stage antigen-1; LSA-1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     The Plasmodium falciparum liver stage antigen LSA-1 is well
            conserved and harbors major B- and T-cell epitopes
  JOURNAL   J. Immunol. (1994) In press
REFERENCE   2  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     Plasmodium falciparum liver stage antigen-1 is well conserved and
            contains potent B and T cell determinants
  JOURNAL   J. Immunol. 153 (1), 190-204 (1994)
  PUBMED    7515922
REFERENCE   3  (bases 1 to 1496)
  AUTHORS   Fidock,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1994) David A Fidock, Laboratory of bio-medical
            parasitology, Pasteur, Institute, 25, rue du Dr. Roux, PARIS, 75724
            CEDEX 15, France
FEATURES             Location/Qualifiers
     source          1. .1496
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="T9/96"
                     /db_xref="taxon:5833"
                     /clone="DG536, DG443, DG750"
                     /cell_type="human erythrocyte culture"
                     /clone_lib="Lambda gt11"
                     /dev_stage="asexual blood stage"
     gene            1. .1482
                     /gene="LSA-1"
     CDS             <1. .1482
                     /gene="LSA-1"
                     /note="nucl. 1 to 1496 corresponds to nucl. 4327 to 5822
                     of published NF54 strain
                     3' extremity of LSA-1 gene of strain T9/96"
                     /codon_start=1
                     /product="liver stage antigen-1"
                     /protein_id="CAA82975.1"
                     /db_xref="GI:510184"
                     /db_xref="UniProt/TrEMBL:Q25886"
                     /translation="QEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDL
                     EQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQRDLEQRKADT
                     KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKE
                     ISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKSADIQNHTLETVNIS
                     DVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIVQYDNFQDEENIGIYKELED
                     LIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDNNFKPNDKSLYDEHIKKYKN
                     DKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKYFMKL"
     repeat_region   1. .639
ORIGIN

  Query Match             100.0%;  Score 1482;  DB 2;  Length 1496;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-158;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
```

```
Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
            |||||||||||||||||||||||||||||||||||||||||||
Db      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

RESULT 4
AR162974
LOCUS AR162974 1482 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 45 from patent US 6270771.
ACCESSION AR162974
VERSION AR162974.1 GI:16233440
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1482)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6270771-A 45 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1482
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 99.9%; Score 1480.4; DB 6; Length 1482;
Best Local Similarity 99.9%; Pred. No. 2e-158;
Matches 1481; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy        61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
```

```
Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

```
RESULT 5
AR178263
LOCUS       AR178263               1482 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 45 from patent US 6319502.
ACCESSION   AR178263
VERSION     AR178263.1  GI:20219401
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 45 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Query Match 99.9%; Score 1480.4; DB 6; Length 1482;
Best Local Similarity 99.9%; Pred. No. 2e-158;
Matches 1481; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
```

```
Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

```
RESULT 6
A28743
LOCUS       A28743                 1493 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA gene 3' (partial).
ACCESSION   A28743
VERSION     A28743.1  GI:1247516
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 32 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
  Query Match           97.6%;  Score 1446;  DB 6;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-154;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
                |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 7
AR162967
LOCUS       AR162967               1493 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 38 from patent US 6270771.
ACCESSION   AR162967
VERSION     AR162967.1  GI:16233432
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 38 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             97.6%;  Score 1446;  DB 6;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-154;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
                |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 8
AR178256
LOCUS       AR178256               1493 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 38 from patent US 6319502.
ACCESSION   AR178256
VERSION     AR178256.1  GI:20219394
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 38 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             97.6%;  Score 1446;  DB 6;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-154;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 9
PFLSA1G
LOCUS       PFLSA1G                5970 bp    DNA     linear   INV 18-APR-2005
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen.
ACCESSION   X56203 S75010 S75012 S75014 S75016
VERSION     X56203.1  GI:9915
KEYWORDS    liver stage antigen.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
```

```
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Zhu,J. and Hollingdale,M.R.
  TITLE     Structure of Plasmodium falciparum liver stage antigen-1
  JOURNAL   Mol. Biochem. Parasitol. 48 (2), 223-226 (1991)
   PUBMED   1840628
REFERENCE   2  (bases 1 to 5970)
  AUTHORS   Zhu,J. and Hollingdale,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1990) Zhu J., Hollingdale M, Biomedical Research
            Institute, 12111 Parklawn Drive, Rockeville, MD 20852, U.S.A
COMMENT     On or before Jun 27, 2005 this sequence version replaced gi:241572,
            gi:241573, gi:241574, gi:241575.
FEATURES             Location/Qualifiers
     source          1. .5970
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="NF54"
                     /db_xref="taxon:5833"
                     /clone="RI-LSA segment, DRAI-LSA segment"
                     /cell_line="NF54 blood stage culture"
                     /clone_lib="pUC13 plasmid, RI-sublib, DRAI-sublib"
                     /dev_stage="liver stage"
     TATA_signal     30. .36
     gene            79. .5808
                     /gene="LSA-1"
     CDS             79. .5808
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage antigen"
                     /protein_id="CAA39663.1"
                     /db_xref="GI:9916"
                     /db_xref="HSSP:1D7M"
                     /db_xref="UniProt/TrEMBL:Q25893"
                     /translation="MKHILYISFYFILVNLLIFHINGKIIKNSEKDEIIKSNLRSGSS
                     NSRNRINEEKHEKKHVLSHNSYEKTKNNENNKFFDKDKELTMSNVKNVSQTNFKSLLR
                     NLGVSENIFLKENKLNKEGKLIEHIINDDDDKKKYIKGQDENRQEDLEEKAAKETLQG
                     QQSDLEQERLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQEQQSDLEQDRLA
                     KEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKL
                     QEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDL
                     EQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLERTKASKETLQEQQSDLEQERLAK
                     EKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQ
                     SDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDR
                     LAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQ
                     EQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLE
                     QERLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKE
                     KLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQS
                     DLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERL
                     AKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQE
                     QQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLERTKASKETLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRA
                     KEKLQEQQSDLEQDRLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDLEQE
                     RRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLANEKLQEQQRDLEQERRAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDL
                     EQERLAKEKLQEQQRDLEQRKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENE
                     RGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGS
                     IKPEQKEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDE
                     FKPIVQYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYE
                     KTKDNNFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSE
                     DITKYFMKL"
ORIGIN

  Query Match           97.3%;  Score 1442;  DB 2;  Length 5970;
  Best Local Similarity 98.3%;  Pred. No. 2.3e-154;

 Matches 1457;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       4327 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 4386

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||||||||||||||||||| |||| ||||||||||||||| ||||| || |||||||||
Db       4387 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 4446

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db       4447 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 4506

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||| |||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db       4507 GAGAGACTTGCTAAAGAAAAGTTGCAAGAGCAACAAAGAGATTTAGAACAAGAGAGACGT 4566

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
Db       4567 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 4626

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||| ||||| || |||||||||||||||||||||||| |||||| ||||||||||||
Db       4627 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAATGAAAAGTTGCAA 4686

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       4687 GAGCAACAAAGAGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 4746

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       4747 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 4806

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4807 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 4866

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||| |||||||||||||||||||||||||||||||| |||||||||||||||||| ||||
Db       4867 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGAGATTTAGAACAAGAGAGACTTGCT 4926

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4927 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 4986

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       4987 AATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTA 5046

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5047 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 5106

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5107 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 5166

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5167 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 5226

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5227 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 5286

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5287 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 5346

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5347 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 5406
```

```
Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5407 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 5466

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5467 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 5526

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5527 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 5586

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5587 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 5646

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5647 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 5706

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5707 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 5766

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       5767 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 5808
```

```
RESULT 10
AE014834
LOCUS       AE014834               253001 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 10 section 6 of 7 of the
            complete sequence.
ACCESSION   AE014834 AE014185
VERSION     AE014834.1  GI:23495173
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253001)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253001)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253001
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
     repeat_region   complement(79. .107)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(113. .136)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   292. .441
                     /rpt_type=tandem
     repeat_region   complement(292. .441)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   455. .479
                     /rpt_type=tandem
     repeat_region   complement(455. .479)
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     gene            <498. .>3257
                     /locus_tag="PF10_0307"
     mRNA            <498. .>3257
                     /locus_tag="PF10_0307"
     CDS             498. .3257
                     /locus_tag="PF10_0307"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN35504.1"
                     /db_xref="GI:23495174"
                     /translation="MASCTGKVNFSSSGKEKGIGLESVLENAFDELISEFKKKFLHKI
                     IKVLEEAKYNENLDIMNERRKNFFLKKYENDQREDEDEYENKIGNVVEKKIEGNVEGN
                     VEGNVEREDIMIEEEKKKKKKKMDSSSSSNDSTFGVYLKSIIDRNKISLLKKLEKQKE
                     LKSVEENNKINYYNNKYKNGIIHYDENKEYHNNLNDKHVYNKNEESMNERTKLVKYTF
                     HDKNKNKNDMYLYPQESHFKYVSERIENLYKISNKLREEKETYHNLKYDKDNENQHFK
                     ENHTEDETDHEQNYEKYNVKLKNYVQNHNTQQMDKYQINEKNNISKLKQYNNNNDNNN
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLKGNMEKNIINNHQKNYISYYNNS
                     FKYPSRSHIFTERFSYHASDEHNEKSDHIYDMSPNKDKNCNNIYGILHNKDKHCNNNY
                     DMSSNKDKHCNNIYDMSPKKDKHCNNIYGILHNKDKHCNNIYGMSPKKDKHCNNIYGI
                     LHNKDKHCNNIYDMSPKKDKHCNNIYEMSPKKDKHCNNIYEMSPKKDKHCNNIYDNSH
                     LSSRGANKLPPSNNIIHEEKKQIINENINIESPQKINDLCNEKNNNLECKKEKIKLEK
                     QFSFDSKDSDDNYVRGKNNMYEQDDLKNNMSQRYYKKVEDFKSYQLNKKISNEKEIIK
                     RPNNNLFFEVIRGKKRKNMQGFECKDCKSFYNEIYSDDDEGKCVHKNFFMKNEQQLKK
                     KININKDVRSKNSSTDDINSVDKMNNVDKMNNANNININNADNINNADNINNADNINNED
                     NINNMDNLNNADNLNNADNLNNADNYPLNKKVHMNNSKKSIYYSHSLSPTCLNAHTDT
                     EKASMNEKKNGNMLNMDNIKKENDLEKEKEKKKKKIIQTFSRHRYHNKINDSPKNFWK
                     FDFFK"
     repeat_region   781. .821
                     /rpt_type=tandem
     repeat_region   841. .870
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   932. .1111
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   1156. .1175
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   complement(1369. .1397)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1439. .1601
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     repeat_region   1474. .1579
                     /rpt_type=tandem
     repeat_region   1479. .1588
                     /rpt_type=tandem
     repeat_region   complement(1618. .1669)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1756. .2184
                     /rpt_type=tandem
     repeat_region   1775. .1954
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     repeat_region   complement(2707. .2735)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2772. .2814
                     /rpt_type=tandem
     repeat_region   2790. .2965
                     /rpt_type=tandem
```

```
repeat_region   2808. .2966
                /rpt_type=tandem
repeat_region   3142. .3171
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(3274. .3347)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3391. .3437)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(3439. .3503)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   3439. .3500
                /rpt_type=tandem
repeat_region   3440. .3500
                /rpt_type=tandem
repeat_region   complement(3561. .3683)
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(3710. .3745)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3763. .3824)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   3763. .3811
                /rpt_type=tandem
repeat_region   complement(3871. .4025)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3922. .3987)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   3926. .3987
                /rpt_type=tandem
repeat_region   complement(4037. .4082)
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   4043. .4082
                /rpt_type=tandem
repeat_region   4043. .4080
                /rpt_type=tandem
repeat_region   complement(4084. .4104)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(4135. .4181)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(4182. .4313)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   4364. .4466
                /rpt_type=tandem
repeat_region   4372. .4457
                /rpt_type=tandem
repeat_region   4377. .4472
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(4512. .4655)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4670. .4775)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   4675. .4768
                /rpt_type=tandem
repeat_region   4870. .4912
                /rpt_type=tandem
repeat_region   complement(4870. .4912)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
```

```
repeat_region   complement(4918. .5000)
                /rpt_type=tandem
                /rpt_unit="(taaaaa)n"
gene            <5162. .>8973
                /locus_tag="PF10_0308"
mRNA            join(<5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .>8973)
                /locus_tag="PF10_0308"
CDS             join(5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .8973)
                /locus_tag="PF10_0308"
                /codon_start=1
                /product="hypothetical protein"
```

```
Query Match             94.9%;  Score 1406.8;  DB 2;  Length 253001;
Best Local Similarity   96.8%;  Pred. No. 4.2e-151;
Matches 1435;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              || |  |||||||||||||||||||||||||||  ||||||||||||| || |  |||
Db     179846 CAGGGGCAACAAAGCGATCTAGAACAAGAGAGACTAGCTAAAGAAAAGTTACAGGGGCAA 179905

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||||||| |||||||||| ||||| |||||||||||||||| | ||| |||||||||
Db     179906 CAAAGCGATCTAGAACAAGAGAGACTAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGAT 179965

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||||||||||||||||||| | |||||||||| ||||||||
Db     179966 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAGCGACAAAGCGATTTAGAACAA 180025

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180026 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTA 180085

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     180086 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGACAGACTTGCTAAAGAA 180145

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||| ||||| || |||||||||||||||||||||||||  ||||| |||||||| ||
Db     180146 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTAGCTAAGGAAAAGTTACAG 180205

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              | || || ||||||||||||| |  ||       |||||||| ||||||||||||||||
Db     180206 GGGCAGCATAGCGATTTAGAACGAACGAAGGCATCTAAAGAAACGTTGCAAGAACAACAA 180265

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180266 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 180325

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180326 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 180385

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180386 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 180445

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180446 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 180505

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     180506 AATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTA 180565

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180566 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 180625

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     180626 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCGAAGGAAATATCTATAATA 180685
```

```
Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180686 GAAAATACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 180745

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180746 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 180805

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180806 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 180865

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180866 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 180925

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180926 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 180985

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180986 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 181045

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181046 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 181105

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181106 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 181165

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181166 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 181225

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181226 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 181285

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db     181286 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 181327


RESULT 11
A28740
LOCUS       A28740                   950 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA-R-NR protein.
ACCESSION   A28740
VERSION     A28740.1  GI:1247512
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 29 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             64.1%;  Score 950;  DB 6;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-98;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 12
AR162961
LOCUS       AR162961                 950 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 32 from patent US 6270771.
ACCESSION   AR162961
VERSION     AR162961.1  GI:16233425
KEYWORDS    .
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 32 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             64.1%;  Score 950;  DB 6;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-98;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy         611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy         671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy         731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy         791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy         851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy         911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 13
AR178250
LOCUS       AR178250                950 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 32 from patent US 6319502.
ACCESSION   AR178250
VERSION     AR178250.1  GI:20219388
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 32 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             64.1%;  Score 950;  DB 6;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-98;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 14
PFALSAC
LOCUS       PFALSAC                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clones KEN1-21b, KEN1-40h, KEN1-56b,
            PNG1-28a, PNG1-28c, BRA1-19e) liver stage-specific antigen 1
            (LSA-1) gene, 3' end of cds.
ACCESSION   L40885
VERSION     L40885.1  GI:725278
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN1-56b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG1-28a) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG1-28c) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: KEN1-21b) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN1-40h) blood stage parasite DNA; and Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA1-19e) blood
            stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN1-56b"
                     /clone="PNG1-28a"
                     /clone="PNG1-28c"
                     /clone="KEN1-21b"
                     /clone="KEN1-40h"
                     /clone="BRA1-19e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59231.1"
                     /db_xref="GI:950664"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       228
                     /gene="LSA-1"
                     /note="clone PNG1-28a"
                     /replace="t"
ORIGIN

  Query Match             56.9%;  Score 843;  DB 2;  Length 843;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-86;
  Matches  843;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

RESULT 15
PFALSAB
LOCUS PFALSAB 843 bp DNA linear INV 19-AUG-1995
DEFINITION Plasmodium flaciparum (clones KEN0-11a,-11b,-37d,-39a,-44e,-51a, -53c,-53d, -56a, PNG0-5h,-7a,-8f,-19a,-26b,-86h,-86i, BRA0-5a) liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION L40884
VERSION L40884.1 GI:725277
KEYWORDS exoerythrocytic; liver stage-specific antigen; liver stage-specific antigen 1; malaria.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 843)
AUTHORS Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M., Hawley,W.A., Collins,W.E. and Lal,A.A.
TITLE Sequence variations in the non-repetitive regions of the liver stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
PUBMED 7477115
COMMENT Original source text: Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-44e) DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-51a) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-7a) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-5h) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-11a) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-11b) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-37d) DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-39a) DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-8f) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-19a) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-26b) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-26b) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-86h) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: PNG0-86i) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-53c) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-53d) blood stage parasite DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone: KEN0-56a) blood stage parasite DNA; and Plasmodium falciparum (individual_isolate Kenyan) (clone: BRA0-5a) blood stage parasite DNA.
Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for Brazil. For each clone, the number following the KEN, PNG or BRA represents group number, the number following the '-' denotes the patient's blood sample number and the lower case letter represents recombinant clones from the corresponding isolates.
FEATURES Location/Qualifiers
source 1. .843
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/isolate="Kenyan"
/db_xref="taxon:5833"
/clone="KEN0-51a"
/clone="KEN0-44e"
/clone="PNG0-7a"
/clone="PNG0-5h"
/clone="KEN0-11a"
/clone="KEN0-11b"
/clone="KEN0-37d"
/clone="KEN0-39a"
/clone="PNG0-8f"
/clone="PNG0-19a"
/clone="PNG0-26b"
/clone="PNG0-86h"
/clone="PNG0-86i"
/clone="KEN0-53c"
/clone="KEN0-53d"
/clone="KEN0-56a"
/clone="BRA0-5a"
/dev_stage="blood stage parasite"
gene 1. .843
/gene="LSA-1"
CDS <1. .843
/gene="LSA-1"
/codon_start=1
/product="liver stage-specific antigen 1"
/protein_id="AAB59230.1"
/db_xref="GI:950663"
/translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY FMKL"
variation 60
/gene="LSA-1"
/note="clone BRA0-5a"
/replace="g"

ORIGIN

Query Match 56.8%; Score 841.4; DB 2; Length 843;
Best Local Similarity 99.9%; Pred. No. 3.5e-86;
Matches 842; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 16
PFALSAG
LOCUS       PFALSAG                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN5-44d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40889
VERSION     L40889.1  GI:725282
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN5-44d) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
            Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN5-44d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41599.1"
                     /db_xref="GI:950668"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPKQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.8%;  Score 841.4;  DB 2;  Length 843;
  Best Local Similarity   99.9%;  Pred. No. 3.5e-86;
  Matches  842;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 17
PFALSAH
LOCUS       PFALSAH                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN6-21a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40890
VERSION     L40890.1  GI:725283
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN6-21a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN6-21a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41600.1"
                     /db_xref="GI:950669"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEYIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.8%;  Score 841.4;  DB 2;  Length 843;
  Best Local Similarity   99.9%;  Pred. No. 3.5e-86;
  Matches  842;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGTATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 18
PFALSAI
LOCUS       PFALSAI                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN7-40c) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40891
VERSION     L40891.1  GI:725284
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN7-40c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
```

```
          Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN7-40c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41601.1"
                     /db_xref="GI:950670"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDY
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match             56.8%;  Score 841.4;  DB 2;  Length 843;
  Best Local Similarity   99.9%;  Pred. No. 3.5e-86;
  Matches  842;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATTATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 19
PFALSAE
LOCUS       PFALSAE                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN3-39d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40887
VERSION     L40887.1  GI:725280
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN3-39d) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN3-39d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41597.1"
                     /db_xref="GI:950666"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDASLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match             56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 5.4e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCTTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 20
PFALSAF
LOCUS       PFALSAF                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN4-32a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40888
VERSION     L40888.1  GI:725281
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
  PUBMED    7477115
COMMENT     Original source text: Plasmodium falciparum (clone: KEN4-32a) blood
            stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /clone="KEN4-32a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41598.1"
                     /db_xref="GI:950667"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKMKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 5.4e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATGAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 21
PFALSAK
LOCUS       PFALSAK                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN9-37e) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40893
VERSION     L40893.1  GI:725286
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN9-37e) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN9-37e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41603.1"
                     /db_xref="GI:950672"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKDENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 5.4e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAGATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 22
PFALSAL
LOCUS       PFALSAL                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone PNG3-5f,-68f,-68h) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40834
VERSION     L40834.1  GI:725287
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-5f) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG3-68h) blood
            stage parasite DNA; and Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-68f) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="PNG3-5f"
                     /clone="PNG3-68h"
                     /clone="PNG3-68f"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59233.1"
                     /db_xref="GI:950673"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEE
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       573
                     /gene="LSA-1"
                     /note="clone PNG3-68h"
                     /replace="c"
ORIGIN

  Query Match           56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity  99.8%;  Pred. No. 5.4e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAGAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 23
PFALSAM
LOCUS       PFALSAM                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clones BRA3-5b,-19a,-14b,-28e) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40835
VERSION     L40835.1  GI:725288
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
```

```
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA3-14b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA3-5b) blood stage
            parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: BRA3-28e) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: BRA3-29b) blood stage parasite
            DNA; and Plasmodium falciparum (individual_isolate Kenyan) (clone:
            BRA3-19a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal 2. The following clones have the sam
            e nucleotide sequence: BRA3-5b; BRA3-19a; BRA3-14b (one silent
            change at 5181,
            A to G) and BRA3-28e (one silent change at 5040, T to C) Clone name
            abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for
            Brazil. For each clone, the number following the KEN, PNG or BRA
            represents group number, the number following the '-' denotes the
            patient's blood sample number and the lower case letter represents
            recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="BRA3-14b"
                     /clone="BRA3-5b"
                     /clone="BRA3-28e"
                     /clone="BRA3-29b"
                     /clone="BRA3-19a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59234.1"
                     /db_xref="GI:950674"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       75
                     /gene="LSA-1"
                     /note="clone BRA3-28e"
                     /replace="c"
     variation       216
                     /gene="LSA-1"
                     /note="clone BRA3-14b"
                     /replace="g"
ORIGIN

  Query Match           56.6%;  Score 838.2;  DB 2;  Length 843;
  Best Local Similarity   99.6%;  Pred. No. 8.1e-86;
  Matches  840;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 24
PFALSA1N
LOCUS       PFALSA1N                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40947
VERSION     L40947.1  GI:725317
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA6-29c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
```

```
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA6-29c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41592.1"
                     /db_xref="GI:725318"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVSDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.5%;  Score 836.6;  DB 2;  Length 843;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-85;
  Matches  839;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAGTGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 25
PFALSAD
LOCUS       PFALSAD                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN2-51b, PNG2-7f,-9f,-11a,-11b,-19d,
            -26e, BRA2-14c,-30b,-30) liver stage-specific antigen 1 (LSA-1)
            gene, 3' end of cds.
ACCESSION   L40886
VERSION     L40886.1  GI:725279
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
  PUBMED    7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN2-51b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-30e) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG2-26e) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-14c) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA2-30b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG2-7f) blood stage
            parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: PNG2-9f) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-11a) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            PNG2-11b) blood stage parasite DNA; and Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-19d) blood stage parasite
            DNA.
            1. Last author is Altaf A. Lal 2. Following clones also have the sa
            me nucleotide sequence: KEN121b, KEN140h, KEN156b, PNG1281 (at 5193
            A to T), P
            NG128c, BRA119e
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN2-51b"
                     /clone="BRA2-30e"
                     /clone="PNG2-26e"
                     /clone="BRA2-14c"
                     /clone="BRA2-30b"
```

```
                     /clone="PNG2-7f"
                     /clone="PNG2-9f"
                     /clone="PNG2-11a"
                     /clone="PNG2-11b"
                     /clone="PNG2-19d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59232.1"
                     /db_xref="GI:950665"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       222
                     /gene="LSA-1"
                     /note="clone PNG2-9f"
                     /replace="c"
     variation       240
                     /gene="LSA-1"
                     /note="clones PNG2-7f, BRA2-30b and BRA2-30e"
                     /replace="g"
ORIGIN

  Query Match             56.5%;  Score 836.6;  DB 2;  Length 843;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-85;
  Matches  839;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 26
PFALSAN
LOCUS       PFALSAN                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA4-28a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40836
VERSION     L40836.1  GI:725289
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA4-28a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="BRA4-28a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41604.1"
                     /db_xref="GI:950675"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEERIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
  Query Match             56.5%;  Score 836.6;  DB 2;  Length 843;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-85;
  Matches  839;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAGAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 27
PFALSAO
LOCUS       PFALSAO                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA5-32f) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40837
VERSION     L40837.1  GI:725290
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (clone: BRA5-32f) blood
            stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /clone="BRA5-32f"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41605.1"
                     /db_xref="GI:950676"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDELDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNGNLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match             56.5%;  Score 836.6;  DB 2;  Length 843;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-85;
  Matches  839;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy         1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy         1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy         1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 ATAGAGAAAAATGGAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy         1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy         1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy         1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy         1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy         1480 TAA 1482
                |||
Db          841 TAA 843
```

```
RESULT 28
PFALSA1C
LOCUS       PFALSA1C                843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA9-32b) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40910
VERSION     L40910.1  GI:725297
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA9-32b) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA9-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41582.1"
                     /db_xref="GI:725298"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
  Query Match           56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity  99.4%;  Pred. No. 1.9e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy          700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
                ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db           61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy          760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy          820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy          880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy          940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy         1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db          361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy         1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy         1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          481 CAATATGACAATTTCCAAGATGAAGAAAGCATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy         1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy         1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy         1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy         1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy         1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840
```

```
Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 29
PFALSAJ
LOCUS       PFALSAJ                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN8-32b) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40892
VERSION     L40892.1  GI:725285
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN8-32b) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN8-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41602.1"
                     /db_xref="GI:950671"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAGCATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 30
PFALSA1A
LOCUS       PFALSA1A                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40908
VERSION     L40908.1  GI:725293
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA7-30a) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
```

```
          and BRA for Brazil. For each clone, the number following the KEN,
          PNG or BRA represents group number, the number following the '-'
          denotes the patient's blood sample number and the lower case letter
          represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA7-30a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59228.1"
                     /db_xref="GI:725294"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFRPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       240
                     /gene="LSA-1"
                     /note="clone BRA7-30a"
                     /replace="g"
ORIGIN

  Query Match             56.2%;  Score 833.4;  DB 2;  Length 843;
  Best Local Similarity   99.3%;  Pred. No. 2.8e-85;
  Matches  837;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAG 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAGGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


Search completed: December 20, 2005, 10:43:01
Job time : 8596.96 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:12:49 ; Search time 331.317 Seconds
(without alignments)
7951.129 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1482	100.0	1482	3	US-08-098-327E-41	Sequence 41, Appl
	2	1482	100.0	1482	3	US-08-462-625-41	Sequence 41, Appl
	3	1480.4	99.9	1482	3	US-08-098-327E-45	Sequence 45, Appl
	4	1480.4	99.9	1482	3	US-08-462-625-45	Sequence 45, Appl
	5	1446	97.6	1493	3	US-08-098-327E-38	Sequence 38, Appl
	6	1446	97.6	1493	3	US-08-462-625-38	Sequence 38, Appl
	7	950	64.1	950	3	US-08-098-327E-32	Sequence 32, Appl
	8	950	64.1	950	3	US-08-462-625-32	Sequence 32, Appl
	9	853.2	57.6	1320	2	US-08-257-073-15	Sequence 15, Appl
	10	472.8	31.9	988	3	US-08-098-327E-34	Sequence 34, Appl
	11	472.8	31.9	988	3	US-08-462-625-34	Sequence 34, Appl
	12	470.8	31.8	954	3	US-08-098-327E-37	Sequence 37, Appl
	13	470.8	31.8	954	3	US-08-462-625-37	Sequence 37, Appl
	14	129.8	8.8	1039	3	US-09-902-540-1280	Sequence 1280, Ap
	15	120.8	8.2	612	3	US-09-902-540-1357	Sequence 1357, Ap
	16	116.2	7.8	3095	9	5231168-1	Patent No. 5231168
	17	111.6	7.5	1827	3	US-09-270-767-1308	Sequence 1308, Ap
	18	111.6	7.5	1827	3	US-09-270-767-16590	Sequence 16590, A
c	19	110.2	7.4	7218	2	US-08-232-463-14	Sequence 14, Appl
	20	101.2	6.8	1956	3	US-08-559-896B-1	Sequence 1, Appli
	21	101.2	6.8	1956	3	US-09-351-794A-1	Sequence 1, Appli
	22	95	6.4	187169	3	US-09-949-016-12776	Sequence 12776, A
	23	95	6.4	191569	3	US-09-949-016-15940	Sequence 15940, A
	24	94.8	6.4	614	3	US-09-902-540-1318	Sequence 1318, Ap
	25	93.6	6.3	5361	3	US-08-973-462-2	Sequence 2, Appli
	26	93.6	6.3	6152	3	US-08-973-462-1	Sequence 1, Appli
c	27	91.6	6.2	29927	3	US-09-949-016-11814	Sequence 11814, A
c	28	91.6	6.2	29927	3	US-09-949-016-17474	Sequence 17474, A
c	29	91.6	6.2	29927	3	US-09-949-016-17475	Sequence 17475, A
c	30	91.6	6.2	37802	3	US-09-949-016-12639	Sequence 12639, A
	31	89.6	6.0	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	32	89.6	6.0	1664976	3	US-09-692-570-1	Sequence 1, Appli
c	33	88.4	6.0	5340	3	US-09-627-122-21	Sequence 21, Appl
	34	87	5.9	699	3	US-09-248-796A-9722	Sequence 9722, Ap
	35	86.8	5.9	4766	6	PCT-US93-07261-10	Sequence 10, Appl
	36	86.4	5.8	601	3	US-09-949-016-203127	Sequence 203127,
	37	86.4	5.8	601	3	US-09-949-016-203169	Sequence 203169,
c	38	85	5.7	147382	3	US-09-949-016-14624	Sequence 14624, A
	39	83.2	5.6	1669	3	US-09-461-697-184	Sequence 184, App
	40	82.8	5.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
	41	81.4	5.5	1685	3	US-10-104-047-1957	Sequence 1957, Ap
	42	81.4	5.5	2223	2	US-08-257-073-4	Sequence 4, Appli
	43	81.2	5.5	696	3	US-09-461-697-193	Sequence 193, App
	44	81.2	5.5	699	3	US-09-461-697-191	Sequence 191, App
	45	81.2	5.5	717	3	US-09-461-697-189	Sequence 189, App

ALIGNMENTS

```
RESULT 1
US-08-098-327E-41
; Sequence 41, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
```

```
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-41

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-246;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 2
US-08-462-625-41
; Sequence 41, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
```

```
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-41

  Query Match           100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-246;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              ||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482

RESULT 3
US-08-098-327E-45
; Sequence 45, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-45

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 3.5e-246;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 4
US-08-462-625-45
; Sequence 45, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-45

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 3.5e-246;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
```

```
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 5
US-08-098-327E-38
; Sequence 38, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-38

  Query Match             97.6%;  Score 1446;  DB 3;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 2.9e-240;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 6
US-08-462-625-38
; Sequence 38, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-38

  Query Match             97.6%;  Score 1446;  DB 3;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 2.9e-240;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197
```

```
Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479
```

RESULT 7
US-08-098-327E-32
; Sequence 32, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; APPLICANT: DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/098,327E
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 92/13884
; PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-32

Query Match 64.1%; Score 950; DB 3; Length 950;
Best Local Similarity 100.0%; Pred. No. 5.8e-155;

Matches 950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950
```

RESULT 8
US-08-462-625-32
; Sequence 32, Application US/08462625
; Patent No. 6319502
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine

```
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-32

  Query Match             64.1%;  Score 950;  DB 3;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-155;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy         611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy         671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy         731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy         791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy         851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy         911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 9
US-08-257-073-15
; Sequence 15, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
```

```
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-15

  Query Match             57.6%;  Score 853.2;  DB 2;  Length 1320;
  Best Local Similarity   87.9%;  Pred. No. 2.7e-138;
  Matches  930;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy        425 ATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 484
              || || || | | |  ||   | ||    ||    |   | | |   | | |   |
Db        263 ATGTAAAAAATGTGTCACAAACAAATTTCAAAAGTCTTTTAAGAAATCTTGGTGTTTCAG 322

Qy        485 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 544
              | | | |     |||||||||       | |  |  || |  | |||    || | | |
Db        323 AGAATATATTCCTTAAAGAAAATAAATTAAATAAGGAAGGGAAATTAATTGAACACATAA 382

Qy        545 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAG 604
                | | ||  |     || || || | |  |     || | || |   |    |||
Db        383 TAAATGATGATGACGATAAAAAAAAATATATTAAAGGGCAAGACGAAAACAGACAAGAAG 442

Qy        605 AAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATT 664
              |       || |   ||    |  |  ||||   ||||||||||||||||||||||||
Db        443 ATCTTGAAGAAAAAGCGCGCGCATCTAAAGAAACGAGGAAGGCTGATACGAAAAAAATT 502

Qy        665 TAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAA 724
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        503 TAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTAGAAA 562

Qy        725 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 622

Qy        785 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAA 844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAA 682

Qy        845 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 742

Qy        905 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 802

Qy        965 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 862

Qy       1025 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 922

Qy       1085 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 982

Qy       1145 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1042

Qy       1205 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1102

Qy       1265 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1162

Qy       1325 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1222

Qy       1385 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1282

Qy       1445 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||
Db       1283 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1320


RESULT 10
US-08-098-327E-34
; Sequence 34, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-34
```

```
  Query Match           31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity  95.7%;  Pred. No. 7.3e-73;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db        661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 11
US-08-462-625-34
; Sequence 34, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-34

  Query Match           31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity  95.7%;  Pred. No. 7.3e-73;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db        661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 12
US-08-098-327E-37
; Sequence 37, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 1.6e-72;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db        629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

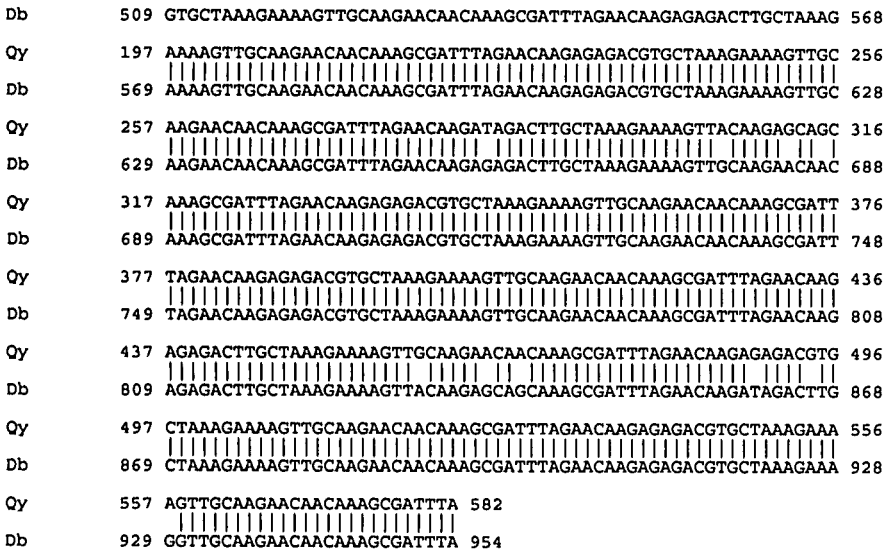
Qy        557 AGTTGCAAGAACAACAAAGCGATTTA 582
              |||||||||||||||||||||||||
Db        929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 13
US-08-462-625-37
; Sequence 37, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 1.6e-72;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

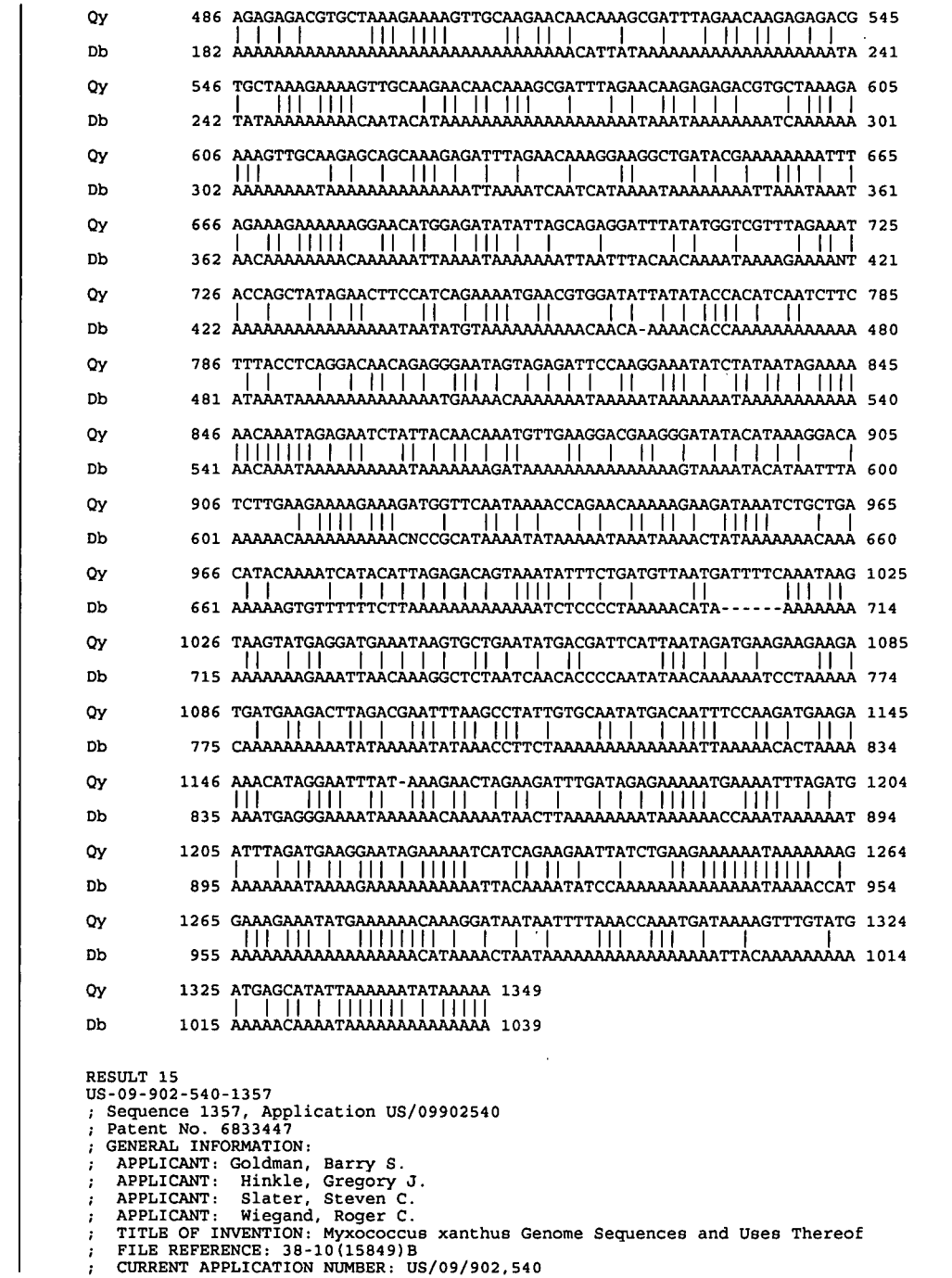
Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
```



RESULT 14
US-09-902-540-1280
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 8.8%; Score 129.8; DB 3; Length 1039;
Best Local Similarity 47.2%; Pred. No. 7.4e-14;
Matches 493; Conservative 0; Mismatches 544; Indels 8; Gaps 3;





RESULT 15
US-09-902-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540

```
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(612)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

  Query Match             8.2%;  Score 120.8;  DB 3;  Length 612;
  Best Local Similarity   50.3%;  Pred. No. 2.5e-12;
  Matches  293;  Conservative    0;  Mismatches  289;  Indels    0;  Gaps    0;

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
              || |  | | |    || |||  |   | || | |||||| ||| | | ||   |  | |
Db          9 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 68

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
              | |||      ||  |  ||   | | | | |||  | |     || |||| ||| |
Db         69 AAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 128

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
               || ||||  | || |||||  | | |||       | | | |||||| | | |  | |
Db        129 AAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA 188

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
              | |  ||| |      |    ||  |     ||| ||  ||  |  |  |  ||| ||
Db        189 AAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAATA 248

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                || |  |  |   |  |||| |  ||  | || || |  || |  |  |  ||
Db        249 AAAAAAAAAAATAAAATAAAAATAAAAAAAAATAAAATATTAAAAAAAAATTAAAAAAAAA 308

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
              ||   ||  |  || |  | ||    || |  |  |||| | |  || | | ||| ||
Db        309 AAAANTAAAATAAAAAAAAAAAAAAAANTAAAAAAATAAAAAAAAAATAAATAAAAAAAAA 368

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              || || |    | | || |   |||    | |  |   | || ||  || | ||||
Db        369 ATAAAAAAAAAAAAATAAAATATAAAAATAAAAAAAAAAAAAAAATAAAAAAAAAAAAAA 428

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                | | || | | |    ||   |||| | || ||  ||| ||| || |||||| |||
Db        429 AAAAAAAAAAATTAAAAAAAAATTAAAAAATAATAAAAAAAAAAAAAATAAAAAAAAAAA 488

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
              | || ||   |||  |||  | ||||     |  |  |  | |  |||||| | |||
Db        489 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 548

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              ||  || |  |    || ||  ||||  |||||   | ||||
Db        549 AAAAATTAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 590


RESULT 16
5231168-1
;Patent No. 5231168
;    APPLICANT: DZIEGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN;
;VUUST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
;    TITLE OF INVENTION: MALARIA ANTIGEN
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/409,658
;      FILING DATE: 18-SEP-1989
;SEQ ID NO:1:
;      LENGTH: 3095
5231168-1

  Query Match             7.8%;  Score 116.2;  DB 9;  Length 3095;
  Best Local Similarity   44.4%;  Pred. No. 1.8e-11;
  Matches  651;  Conservative    0;  Mismatches  803;  Indels   13;  Gaps    4;

Qy         17 ATCTAGAACAAGAGAGACGTGCTAAAGAAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 76
              || |  | || || | |    |  ||||  | |  |    | || |      |   |||
Db       1133 ATATTCAGCATGAAATAGTAGAAGTAGAAGAAATACTTCCAGAAGATGATAAAAATGAAA 1192

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              ||| |  || ||  | || | |  |  ||||    | |  ||    || | || | | |
Db       1193 AAGTTGAACATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAG 1252

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |    | ||||    |  | |   || |||      | |||| | || | |  ||  ||||
Db       1253 GTCAACATGAAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATGATAAAAATGAAAAAG 1312

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              |     ||  |  | ||   ||   |  | |     |  |   || |||||     |
Db       1313 TTGAACATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTC 1372

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              || |  ||  |   ||  | ||| ||  |  ||  |   |  ||||   | |||   | |
Db       1373 AACATGAAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTGAAC 1432

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              |          ||||  ||   |  |     |||| |     | ||| ||        ||
Db       1433 ATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTCAACATG 1492

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              |  |  ||||      | | ||||||   | | |||| |  |||  ||   ||  |||
Db       1493 AAATAGTAGAG------GTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAAC 1546

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |  | |  |||  |          || | ||    |  ||| |||     || ||
Db       1547 ATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTCAACATG 1606

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              | || | || |  ||||   | |   ||  |||  || |   |   | | || |
Db       1607 AAATAGTAGAGGTTGAAGAAATTCTACCAGAAGAAGATAAAAATGAAAAAGGTCAACATG 1666

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 616
              |  |  |||     ||      | ||  | |||| | |  ||  ||||   |     ||
Db       1667 AAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAACATGAAA 1726

Qy        617 AGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAA 676
              |     ||  |  | |   | | | | |  |||||  |  | |  || |
Db       1727 TAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAACATGAAATAG 1786

Qy        677 AGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAG 736
              ||   || ||| ||  |  ||||  | |  |     |  |  |  ||| |  | ||  | |
Db       1787 TAGAGGTTGAAGAAATTCTTCCAGAAATTGTTGAAATTGAAGAAGTACCATCACAAACAA 1846

Qy        737 AACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGG 796
              |          | | | ||||  |  | |      |  | |  |    ||   |   |
Db       1847 ATAACAATGAAAATATTGAAACTATAAAACCAGAAGAAAAAAAGAATGAATTTAGTGTTG 1906

Qy        797 ACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAG 856
              |  | | ||  | |     |||  ||  ||   || ||| | || |  |||   | |  |
Db       1907 AAGAAAAAGCAATTCCACAAGAACCCGTGG---TACCTACATTAAATGAAAATGAAAACG 1963

Qy        857 AATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAA 916
              ||   | | ||  ||  |  |   |   | |||  ||  || || ||  |  |
Db       1964 TTACTCCCAAACCATCTGAAGGTGAATCCACTAAACCAGATATAGTTCAAATTAAAATAG 2023

Qy        917 AGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATC 976
              ||||   |  |  |||  | |  ||| |  ||    |  || |   | |||
Db       2024 TACAAGAAAATAAACCAAATAAAAAGGAAACACCAGTAGTAGATGGTCCAAAACATGTAG 2083

Qy        977 ATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGG 1036
              |  |       || | ||  |  |||||  | |||  | |    |  | | ||| |
Db       2084 AACAAAATATACAAGAAGATGATAATGATGAAGAGGATGATGATGATATAGATTTTGAAG 2143

Qy       1037 ATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACT 1096
```

```
Db       2144 GATTATCAAGAAAAGATGATGAAAAGGATTCATCAAATAAAAATAAAAAGAAATCATCTT 2203

Qy       1097 TAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAA 1156

Db       2204 TTATAACATATATATCTACAAAGAAATTTAAAAAAGTATCTCAAACTATTGTAAGTGTTA 2263

Qy       1157 TTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAG 1216

Db       2264 TGATTAATGCATATGATGGTGTTATTCAAGTTGTAAGTACAATTAAAGGAATAGCAAAGG 2323

Qy       1217 GAATAGAAAAATCAT--CAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATA 1274

Db       2324 ATATAGTAATATTTTTCCAAAACATTTAAATAATTAACAAAAAAAAAAAAAAAAAAAAAA 2383

Qy       1275 TGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATAT 1334

Db       2384 AAAAAAAATATTAAATAAAATTTTTTTTCTTATTATATGTAACTAATATTATTTATTAAT 2443

Qy       1335 TAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATT 1394

Db       2444 AAATATATATATTATATAAAAGAATACTAGGATTTCTGTATATATAGAAATAAATTCATT 2503

Qy       1395 GTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGA 1454

Db       2504 GTATATTATTATAGAAAAAAATAA--AAATAAATAATTTCTTTTCTTAGTTGTATTAATA 2561

Qy       1455 TATAACTAAATATTTTATGAAACTATA 1481

Db       2562 TTCTAAAAATATAATAAATAAAATATA 2588


RESULT 17
US-09-270-767-1308
; Sequence 1308, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1308
;   LENGTH: 1827
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-1308

  Query Match             7.5%;  Score 111.6;  DB 3;  Length 1827;
  Best Local Similarity   47.8%;  Pred. No. 1.1e-10;
  Matches  454;  Conservative    0;  Mismatches  484;  Indels   12;  Gaps    4;

Qy         24 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAG 83

Db        804 AAAAAAAGAAGATGAAATAAACAAGAATCAAGGGAAACCACGAATAATGGAAAAGGAAAG 863

Qy         84 ACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAA 143

Db        864 A--AGTAAAGAAGAGGAAAAAAGAAAAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAAA 921

Qy        144 AGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 203

Db        922 AGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAA 981

Qy        204 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 263

Db        982 GAGGAAAGAAAAAGAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGAAAAA----AAGAAGA 1037

Qy        264 ACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGA 323

Db       1038 GGAAAGAGAAAGAGAAGAGGAAAGAA-----AGAGAGAACATAATAGGAAAAAAGAAGAG 1092

Qy        324 TTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACA 383

Db       1093 GAAAGAAAAAGAGAAGAGAAAAGAAGAAAAGAAGAGGAAAAACGAAAGGAAGAGGAAAGA 1152

Qy        384 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACT 443

Db       1153 AGAAAGGAAGAGGAAAGGAAAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGGAAAGAGGAG 1212

Qy        444 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGA 503

Db       1213 GAAAGAAGAAAGGAAGAGGAAAGAAGAAAAGAGAAAAGAAGAAGAGATGAGAAGAGAAGA 1272

Qy        504 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 563

Db       1273 AGAGAAGAGGAAAAGAGAAAGGAAGAGGAAAGGAAAGAAGAAGAAAGAAGAGAGGAAGCG 1332

Qy        564 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCA 623

Db       1333 GAAAGGAAAGAAGAGGAAAGGAAAGCAGAGGAAAGAAGAAAGAAAGAGGAAAGAAGAAGA 1392

Qy        624 AAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACA 683

Db       1393 GA-AGAGAAAAGAAGAAGGGAAGAGAAAAGAAGAAGGGAAGAAGAGGAAAGAAGGAAAGA 1451

Qy        684 TGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCC 743

Db       1452 AGAGGAAAGAAGAGAAGAGGAGGAAAAAAGAAAAGAAGAGGAAAGAAGGAAAGATGAGGA 1511

Qy        744 ATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAG 803

Db       1512 AAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAG 1571

Qy        804 AGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTAT 863

Db       1572 AGAAGAGGGAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAGAGAAGAGGA 1631

Qy        864 TACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGA 923

Db       1632 AAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAGACGAGGAACGAAGAAGAGA 1691

Qy        924 TGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAA 973

Db       1692 AGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAA 1741

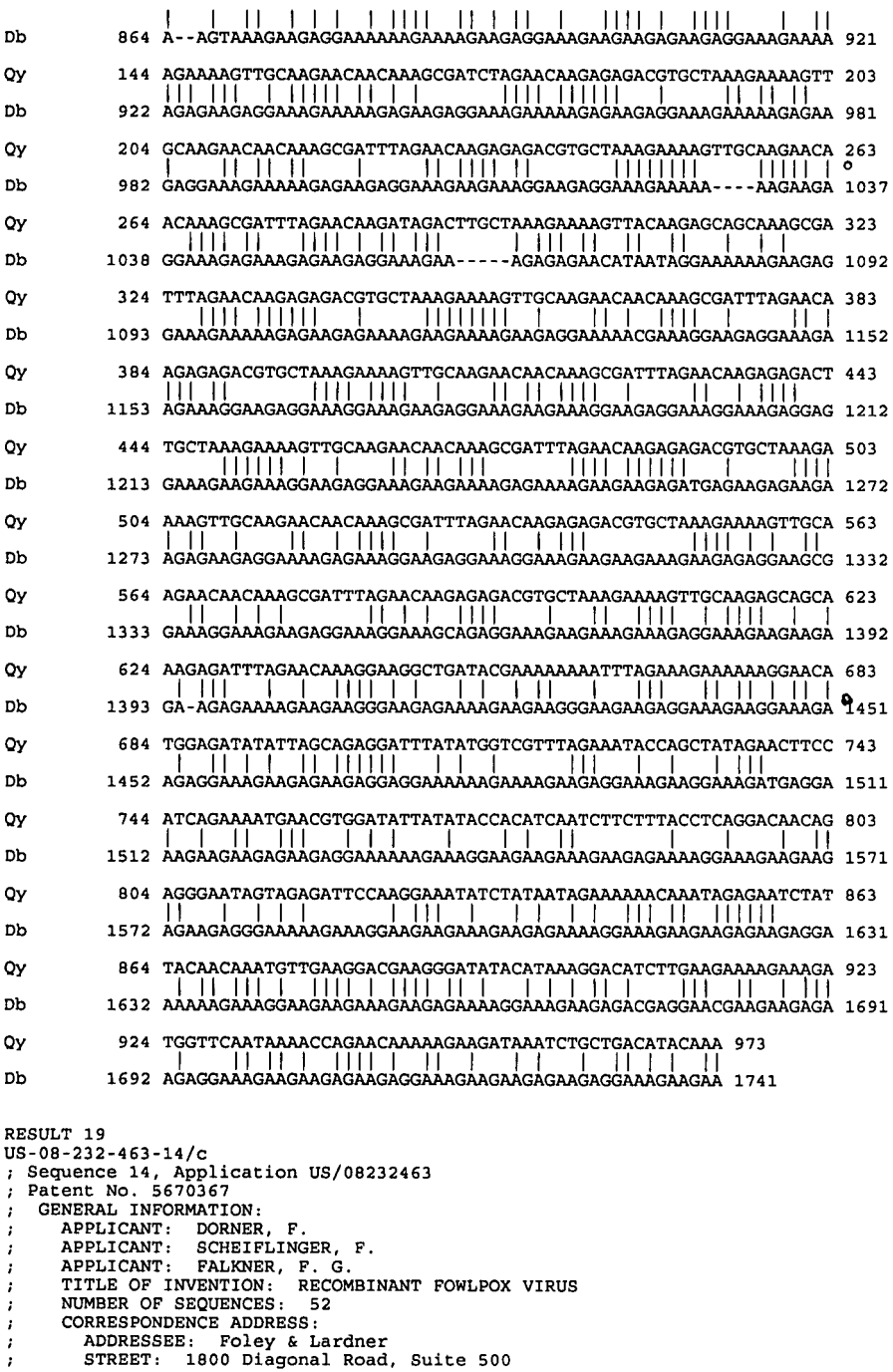

RESULT 18
US-09-270-767-16590
; Sequence 16590, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16590
;   LENGTH: 1827
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-16590

  Query Match             7.5%;  Score 111.6;  DB 3;  Length 1827;
  Best Local Similarity   47.8%;  Pred. No. 1.1e-10;
  Matches  454;  Conservative    0;  Mismatches  484;  Indels   12;  Gaps    4;

Qy         24 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAG 83

Db        804 AAAAAAAGAAGATGAAATAAACAAGAATCAAGGGAAACCACGAATAATGGAAAAGGAAAG 863

Qy         84 ACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAA 143
```



```
Db        864 A--AGTAAAGAAGAGGAAAAAAGAAAAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAAA 921

Qy        144 AGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 203

Db        922 AGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAAAAGAGAA 981

Qy        204 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 263

Db        982 GAGGAAAGAAAAAGAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGAAAAA----AAGAAGA 1037

Qy        264 ACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGA 323

Db       1038 GGAAAGAGAAAGAGAAGAGGAAAGAA-----AGAGAGAACATAATAGGAAAAAAGAAGAG 1092

Qy        324 TTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACA 383

Db       1093 GAAAGAAAAAGAGAAGAGAAAAGAAGAAAAGAAGAGGAAAAACGAAAGGAAGAGGAAAGA 1152

Qy        384 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACT 443

Db       1153 AGAAAGGAAGAGGAAAGGAAAGAAGAGGAAAGAAGAAAGGAAGAGGAAAGGAAAGAGGAG 1212

Qy        444 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGA 503

Db       1213 GAAAGAAGAAAGGAAGAGGAAAGAAGAAAAGAGAAAAGAAGAAGAGATGAGAAGAGAAGA 1272

Qy        504 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 563

Db       1273 AGAGAAGAGGAAAAGAGAAAGGAAGAGGAAAGGAAAGAAGAAGAAAGAAGAGAGGAAGCG 1332

Qy        564 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCA 623

Db       1333 GAAAGGAAAGAAGAGGAAAGGAAAGCAGAGGAAAGAAGAAAGAAAGAGGAAAGAAGAAGA 1392

Qy        624 AAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACA 683

Db       1393 GA-AGAGAAAAGAAGAAGGGAAGAGAAAAGAAGAAGGGAAGAAGAGGAAAGAAGGAAAGA 1451

Qy        684 TGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCC 743

Db       1452 AGAGGAAAGAAGAGAAGAGGAGGAAAAAAGAAAAGAAGAGGAAAGAAGGAAAGATGAGGA 1511

Qy        744 ATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAG 803

Db       1512 AAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAG 1571

Qy        804 AGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTAT 863

Db       1572 AGAAGAGGGAAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAGAGAAGAGGA 1631

Qy        864 TACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGA 923

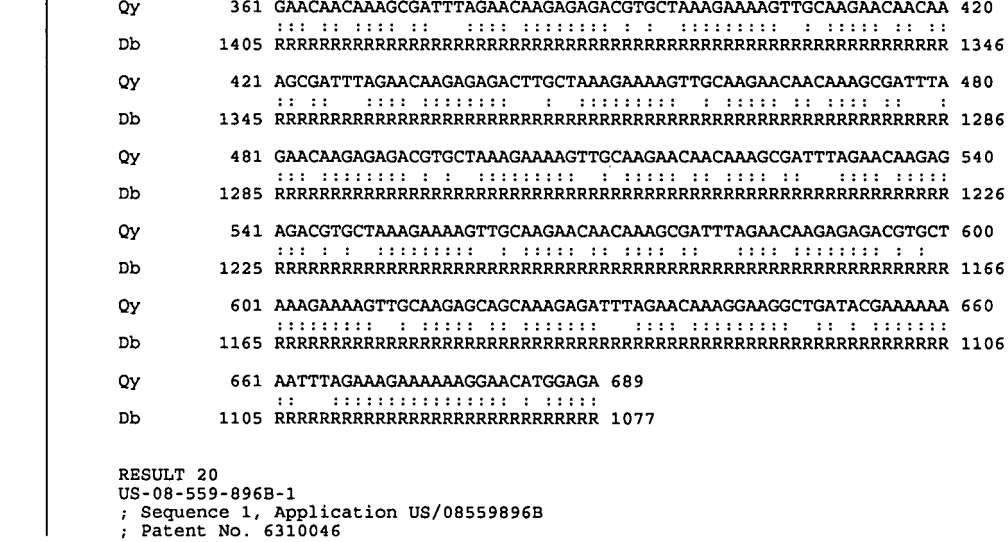
Db       1632 AAAAAGAAAGGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAGACGAGGAACGAAGAAGAGA 1691

Qy        924 TGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAA 973

Db       1692 AGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAA 1741
```

RESULT 19
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 7.4%; Score 110.2; DB 2; Length 7218;
Best Local Similarity 4.1%; Pred. No. 2.1e-10;
Matches 16; Conservative 265; Mismatches 108; Indels 0; Gaps 0;



```
Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Db       1465 AAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1406

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Db       1405 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1346

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Db       1345 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1286

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Db       1285 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1226

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Db       1225 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1166

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Db       1165 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1106

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGA 689

Db       1105 RRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1077
```

RESULT 20
US-08-559-896B-1
; Sequence 1, Application US/08559896B
; Patent No. 6310046

```
;  GENERAL INFORMATION:
;    APPLICANT:  Patrick E. Duffy
;    APPLICANT:  Christian F. Ockenhouse
;    TITLE OF INVENTION:  SEQUESTRIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John Moran
;      STREET:  USA MRMC -  MCMR-JA
;      CITY:  FORT DETRICK, FREDERICK
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21702-5012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,896B
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Moran, John
;      REGISTRATION NUMBER:  26,313
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-7714
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1956 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
US-08-559-896B-1

  Query Match             6.8%;  Score 101.2;  DB 3;  Length 1956;
  Best Local Similarity   43.7%;  Pred. No. 6.5e-09;
  Matches  494;  Conservative    0;  Mismatches  633;  Indels    3;  Gaps    1;

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 406
              | |||| ||    |||| |   ||       | |||||||| | |  |   ||    ||
Db        293 ATGAAATGTATAGAGAAGAGTTAAATAGGATTGAACAAGAAAAAATTAGAAATATGAATA 352

Qy        407 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAAC 466
                |||||  | | |    ||   | | || | |  |  ||   | |     ||||
Db        353 AAGAAGAAATAAATAAAACTTACAAGGACGAAATAAATAATATGAATAGTGATCAAGTTG 412

Qy        467 AACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 526
              |  |||   ||  |||| ||      |        || || |      | |  ||
Db        413 ATAAAATACATAGAGAAGAATTAGAAAAAATCGAAAAAGAAAAAATAAACAAAATGGATA 472

Qy        527 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 586
              |    |||  ||| | |    || |||| | ||  |  ||    |  | |||  |
Db        473 AAGATGAAATAGATAAAATATATAGAGAAGAATTAGACAAAATGGATCGAGATGCAATTT 532

Qy        587 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGG 646
              |    |    |    |||| |       |  |  |   || ||| |||    |||   |
Db        533 ATAGTATGTATATAGAAGATATAAGTAACAAAAATATAAAAGACTTAATTAAAAATGAAA 592

Qy        647 CTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706
                || || || ||  ||  | | | ||||||  ||    |     |     |   | |
Db        593 AGGAAACAAATAAGGATAAAAACAAAAAAAAAGATATAGATATAAACAAAAAGAAAAAAA 652

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766
              | ||    | ||||  ||          ||  |   |    || || |    |  | | |
Db        653 AAGATATAGATATAGATGTAGACATAGATAAAGATATACATAAAGATCATGTAGAAGAAT 712

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826
                  | |   |        || |    |  |   |    |||| |  ||     | |
Db        713 TATACGGAGAAGTAAAAAACAAACTTAGCAAAGAAGAATTAGATAGAATGGACAGAGACG 772

Qy        827 AAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886
              | ||| || | ||  |    |  | |   |||     || |  | | | |    |    ||
Db        773 CATTATATAGAGTATACCTAGAAGAATTAGATAGAATGAACAGAGACGAATTATATAGAG 832

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946
                |  ||  |  ||  | |       ||| ||| ||||      ||  ||  |    | |||
Db        833 TATACCTAGAAGAATTAGAAAAAATAGACAAGGAAGAAAAAGAAAAATTCATAGAGAAA 892

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006
              ||  | | |||  ||   |      |||   ||| | | || | ||   | ||   | |
Db        893 AATTACACAAAATTGAAAAAGAGAAAATAAATAAAATGGATAAAGATCAAATAGATAAAA 952

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066
              |  ||||      || |||  ||     ||   |||   ||     | || |||       |
Db        953 TATATGAAGAAGAATTAAACAAAATGGATAGTGATGAAA---TTCAACATGTAAGGAGAG 1009

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126
              ||||    ||||   | |  | |||  |  || |  | ||  | |    |       ||
Db       1010 CAATATTACAAGATATACAAAAAGAAAAAATACAAAATTTAGAACTAGAAGAAATCGATA 1069

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186
                ||   |||| |||  | | | |     |    |||  |    | ||  |      ||
Db       1070 GACTTTATAAAGAAGAATTAGATAGAATGGATAGAGAAGCAAGGTATGAAATCCCCATGA 1129

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246
              ||||  ||    | ||||   ||||  |    ||| | | |       || |  ||||
Db       1130 GAAATTTAAGCAGAAATGAAAAAGATAATATTATACATAGAAATATTAAAAATGAATCTA 1189

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306
              | ||||||  || |||| |||   | ||| |   | ||   |  | |||  | |   |
Db       1190 ACCAAAAAAATAAGAAAGAAAATGTAAATGTATTTATAATACACGACAATAATGATAGTA 1249

Qy       1307 ATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATA 1366
              || |||| | |    || || |      |   ||  ||  |||| |||||||    ||||
Db       1250 ATAATAATAATAATAATAATAATAGAGATGTGAATAATTTAAATAATAAGCACACAAATA 1309

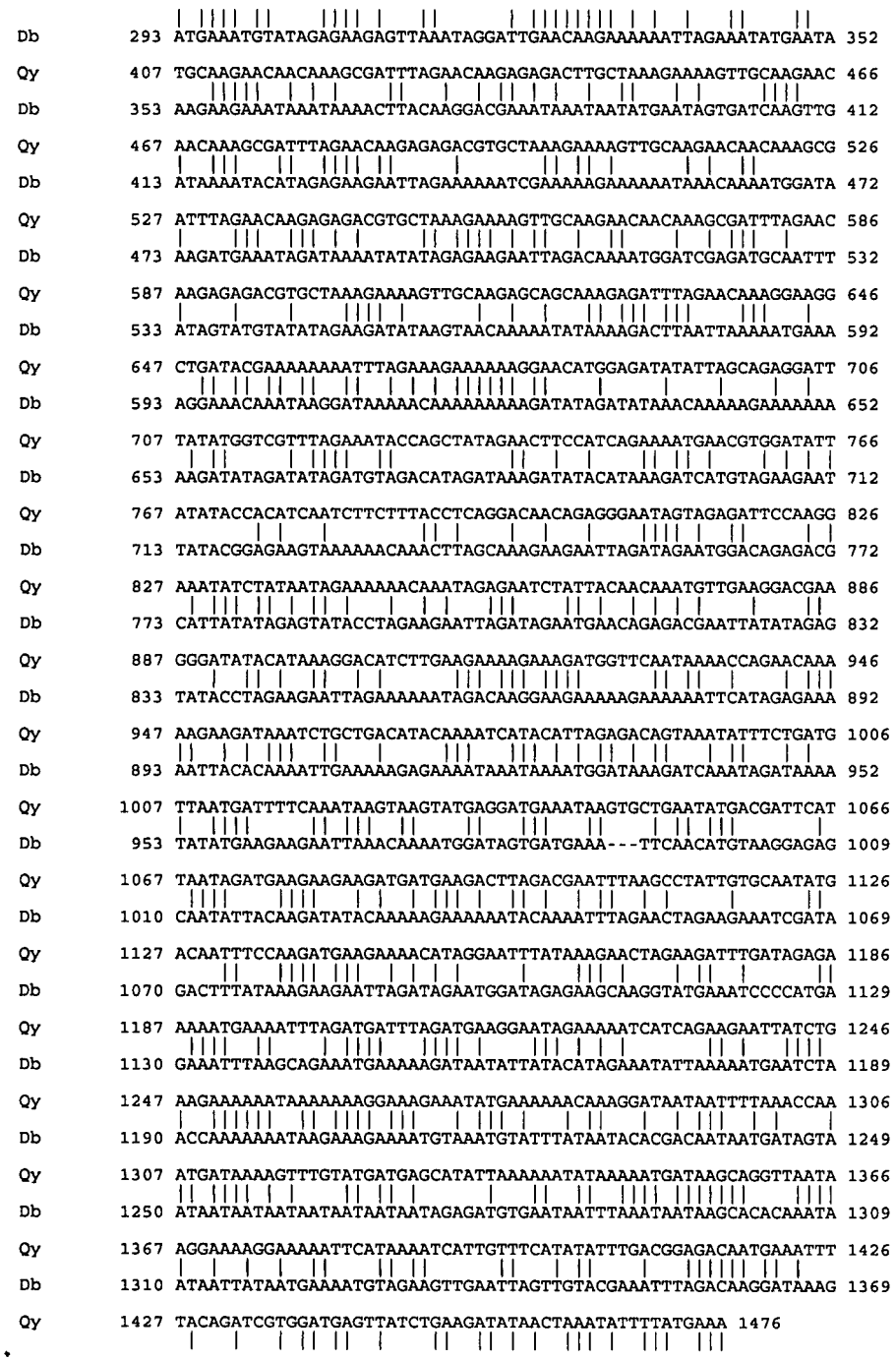
Qy       1367 AGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTT 1426
              |  |  |  |  ||    || ||       ||   | ||    |   |||||| || |
Db       1310 ATAATTATAATGAAAATGTAGAAGTTGAATTAGTTGTACGAAATTTAGACAAGGATAAAG 1369

Qy       1427 TACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAA 1476
              |   |  | || ||  |   ||  || | |  ||| |  |||  |||
Db       1370 GAGCCAAGATAGAAGATATTATAGACTATTTTAACAAAGAAATTAAAAAA 1419


RESULT 21
US-09-351-794A-1
; Sequence 1, Application US/09351794A
; Patent No. 6641815
; GENERAL INFORMATION:
;  APPLICANT: DUFFY, PATRICK E.
;  APPLICANT:  OCKENHOUSE, CHRISTIAN F.
;  TITLE OF INVENTION: SEQUESTRIN
;  FILE REFERENCE: 38644-175519
;  CURRENT APPLICATION NUMBER: US/09/351,794A
;  CURRENT FILING DATE:  1999-07-13
;  PRIOR APPLICATION NUMBER: 08/559,896
;  PRIOR FILING DATE: 1995-11-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1956
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-351-794A-1

  Query Match             6.8%;  Score 101.2;  DB 3;  Length 1956;
  Best Local Similarity   43.7%;  Pred. No. 6.5e-09;
  Matches  494;  Conservative    0;  Mismatches  633;  Indels    3;  Gaps    1;

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 406
```



```
Db        293 ATGAAATGTATAGAGAAGAGTTAAATAGGATTGAACAAGAAAAAATTAGAAATATGAATA 352

Qy        407 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAAC 466

Db        353 AAGAAGAAATAAATAAAACTTACAAGGACGAAATAAATAATATGAATAGTGATCAAGTTG 412

Qy        467 AACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 526

Db        413 ATAAAATACATAGAGAAGAATTAGAAAAAATCGAAAAAGAAAAAATAAACAAAATGGATA 472

Qy        527 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 586

Db        473 AAGATGAAATAGATAAAATATATAGAGAAGAATTAGACAAAATGGATCGAGATGCAATTT 532

Qy        587 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGG 646

Db        533 ATAGTATGTATATAGAAGATATAAGTAACAAAAATATAAAAGACTTAATTAAAAATGAAA 592

Qy        647 CTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706

Db        593 AGGAAACAAATAAGGATAAAAACAAAAAAAAAGATATAGATATAAACAAAAAGAAAAAAA 652

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766

Db        653 AAGATATAGATATAGATGTAGACATAGATAAAGATATACATAAAGATCATGTAGAAGAAT 712

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826

Db        713 TATACGGAGAAGTAAAAAACAAACTTAGCAAAGAAGAATTAGATAGAATGGACAGAGACG 772

Qy        827 AAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886

Db        773 CATTATATAGAGTATACCTAGAAGAATTAGATAGAATGAACAGAGACGAATTATATAGAG 832

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946

Db        833 TATACCTAGAAGAATTAGAAAAAATAGACAAGGAAGAAAAAGAAAAAATTCATAGAGAAA 892

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006

Db        893 AATTACACAAAATTGAAAAAGAGAAAATAAATAAAATGGATAAAGATCAAATAGATAAAA 952

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066

Db        953 TATATGAAGAAGAATTAAACAAAATGGATAGTGATGAAA---TTCAACATGTAAGGAGAG 1009

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126

Db       1010 CAATATTACAAGATATACAAAAAGAAAAAATACAAAATTTAGAACTAGAAGAAATCGATA 1069

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186

Db       1070 GACTTTATAAAGAAGAATTAGATAGAATGGATAGAGAAGCAAGGTATGAAATCCCCATGA 1129

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246

Db       1130 GAAATTTAAGCAGAAATGAAAAAGATAATATTATACATAGAAATATTAAAAATGAATCTA 1189

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306

Db       1190 ACCAAAAAAATAAGAAAGAAAATGTAAATGTATTTATAATACACGACAATAATGATAGTA 1249

Qy       1307 ATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATA 1366

Db       1250 ATAATAATAATAATAATAATAATAGAGATGTGAATAATTTAAATAATAAGCACACAAATA 1309

Qy       1367 AGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTT 1426

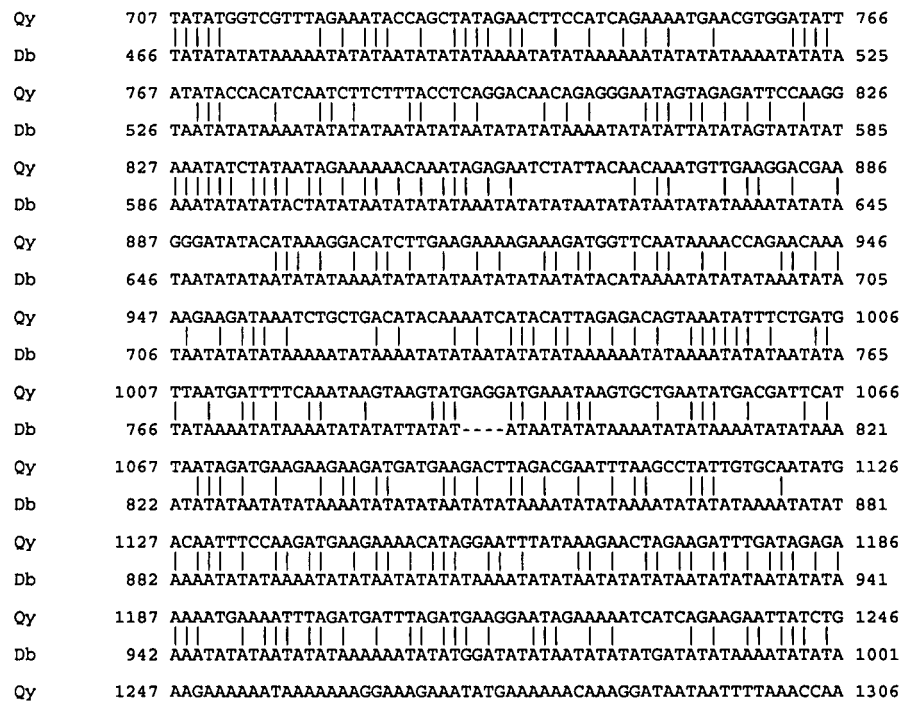
Db       1310 ATAATTATAATGAAAATGTAGAAGTTGAATTAGTTGTACGAAATTTAGACAAGGATAAAG 1369

Qy       1427 TACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAA 1476

Db       1370 GAGCCAAGATAGAAGATATTATAGACTATTTTAACAAAGAAATTAAAAAA 1419
```

RESULT 22
US-09-949-016-12776
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match 6.4%; Score 95; DB 3; Length 187169;
Best Local Similarity 47.0%; Pred. No. 1.1e-07;
Matches 364; Conservative 0; Mismatches 405; Indels 6; Gaps 2;



```
Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766

Db        466 TATATATATAAAAATATATAATATATATAAAATATATAAAAAATATATATAAAATATATA 525

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826

Db        526 TAATATATAAAATATATATAATATATAATATATATAAAATATATATTATATAGTATATAT 585

Qy        827 AAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886

Db        586 AAATATATATACTATATAATATATATAAATATATATAATATATAATATATAAAATATATA 645

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946

Db        646 TAATATATAATATATAAAATATATATAATATATAATATACATAAAATATATATAAATATA 705

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006

Db        706 TAATATATATAAAAATATAAAATATATAATATATATAAAAAATATAAAATATATAATATA 765

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066

Db        766 TATAAAATATAAAATATATATTATAT----ATAATATATAAAATATATAAAATATATAAA 821

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126

Db        822 ATATATAATATATAAAATATATATAATATATAAAATATATAAAATATATATAAAATATAT 881

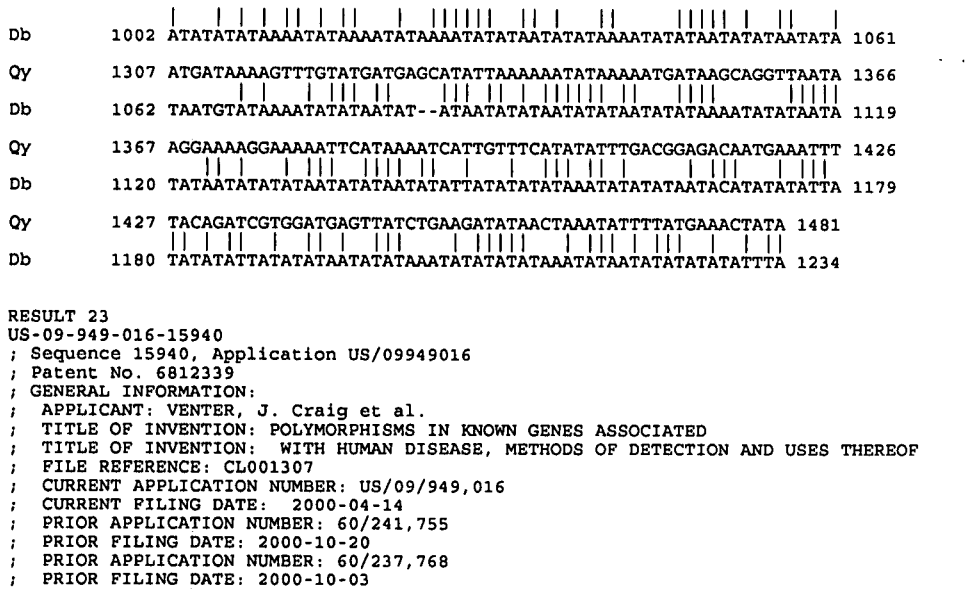
Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186

Db        882 AAAATATATAAAATATATAATATATATAAAATATATAATATATATAATATATAATATATA 941

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246

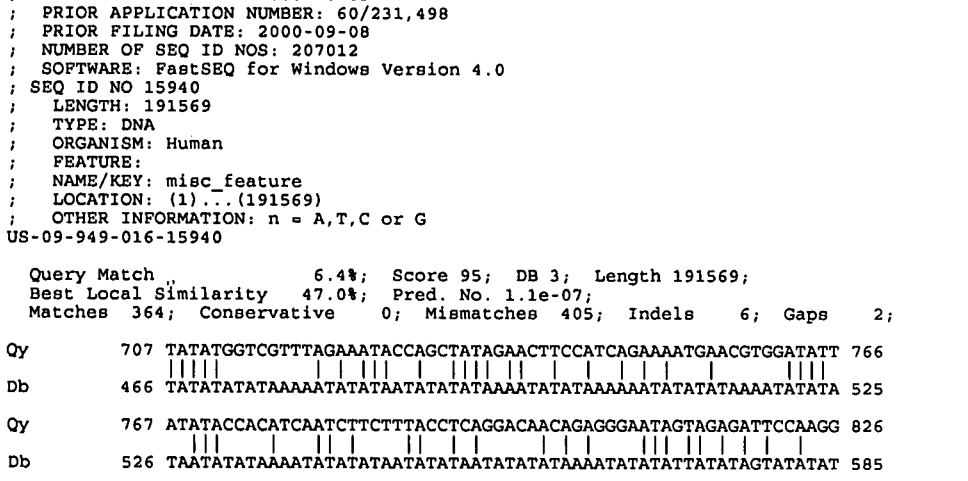
Db        942 AAATATATAATATATAAAAAATATATGGATATATAATATATATGATATATAAAATATATA 1001

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306
```



RESULT 23
US-09-949-016-15940
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

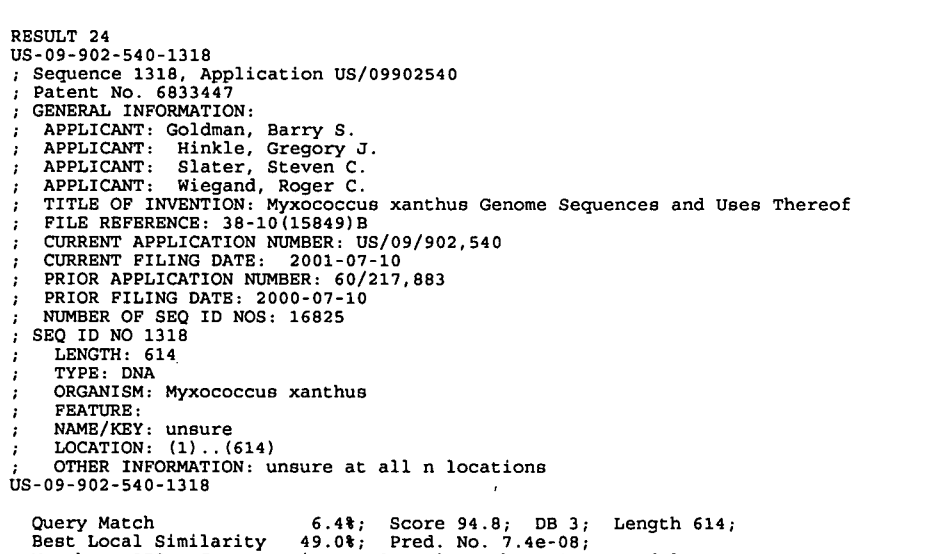
Query Match 6.4%; Score 95; DB 3; Length 191569;
Best Local Similarity 47.0%; Pred. No. 1.1e-07;
Matches 364; Conservative 0; Mismatches 405; Indels 6; Gaps 2;

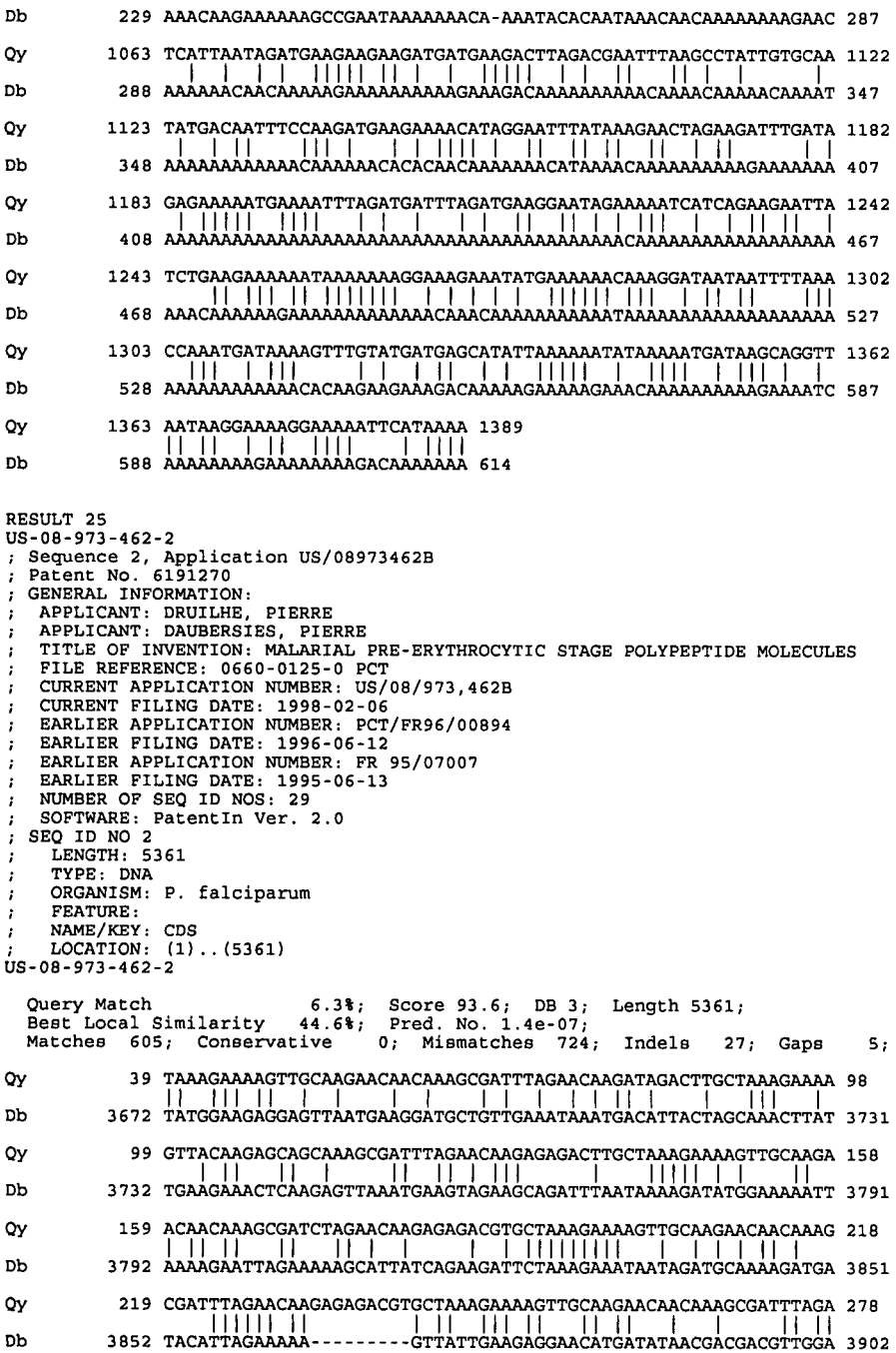




RESULT 24
US-09-902-540-1318
; Sequence 1318, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1318
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(614)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1318

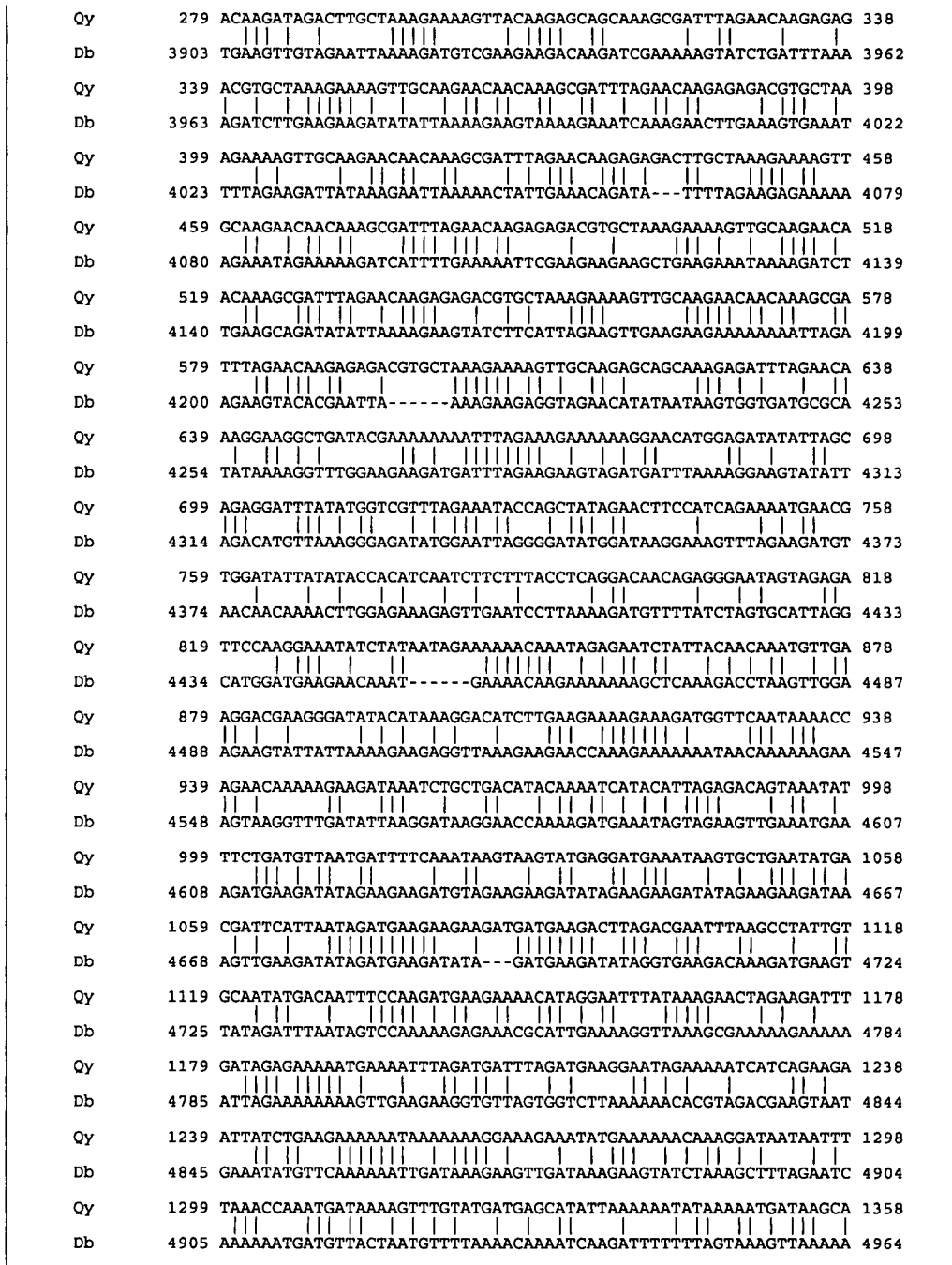
Query Match 6.4%; Score 94.8; DB 3; Length 614;
Best Local Similarity 49.0%; Pred. No. 7.4e-08;
Matches 278; Conservative 0; Mismatches 288; Indels 1; Gaps 1;





RESULT 25
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 6.3%; Score 93.6; DB 3; Length 5361;
Best Local Similarity 44.6%; Pred. No. 1.4e-07;
Matches 605; Conservative 0; Mismatches 724; Indels 27; Gaps 5;

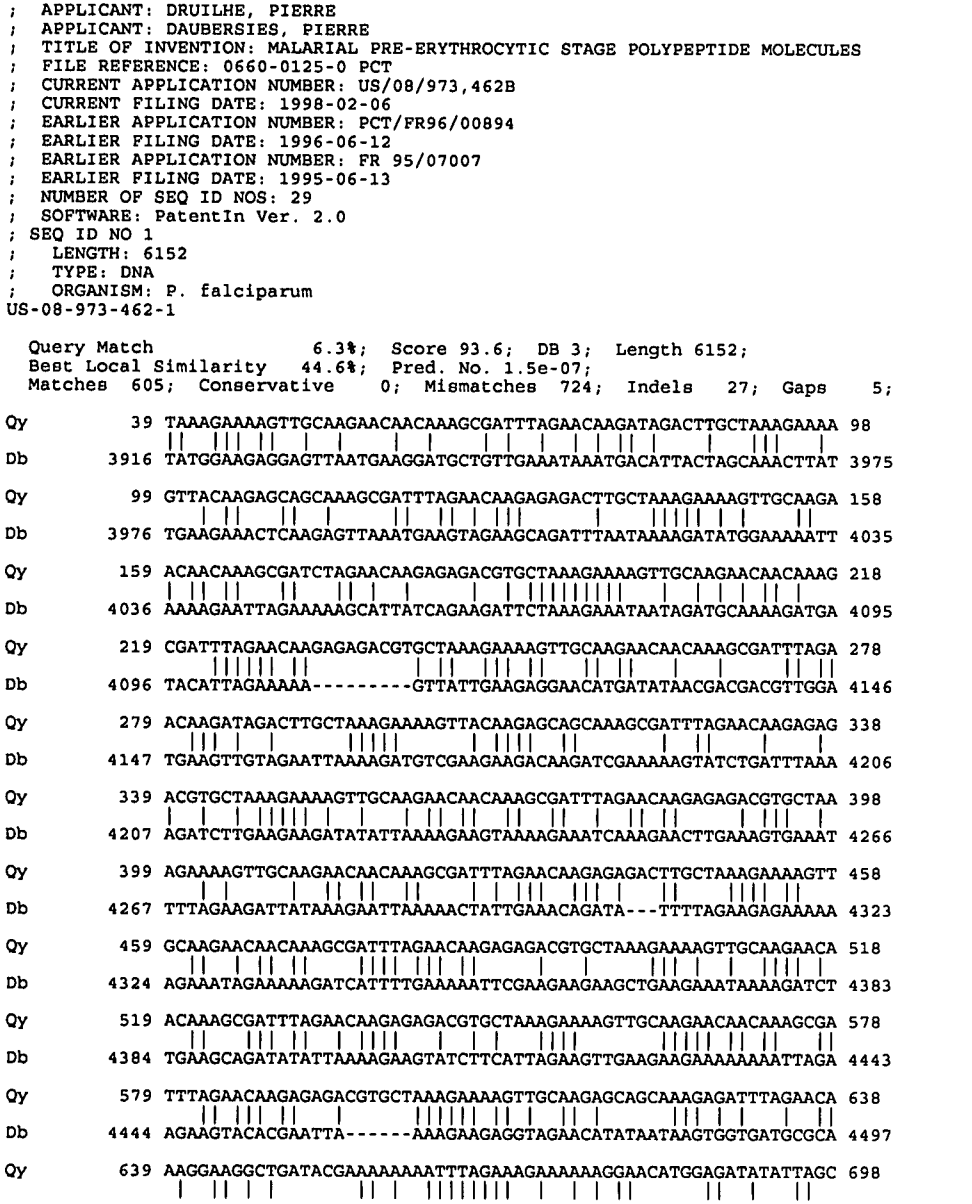




```
Db       4965 CTTCGTAAAAAAATATAAAGTATTTGCTGCACCATT 5000
```

```
RESULT 26
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US-08-973-462-1

  Query Match             6.3%;  Score 93.6;  DB 3;  Length 6152;
  Best Local Similarity   44.6%;  Pred. No. 1.5e-07;
  Matches  605;  Conservative    0;  Mismatches  724;  Indels   27;  Gaps    5;
```



```
Qy         39 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 98
Db       3916 TATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTAT 3975

Qy         99 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGA 158
Db       3976 TGAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATT 4035

Qy        159 ACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 218
Db       4036 AAAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGA 4095

Qy        219 CGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 278
Db       4096 TACATTAGAAAAA---------GTTATTGAAGAGGAACATGATATAACGACGACGTTGGA 4146

Qy        279 ACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAG 338
Db       4147 TGAAGTTGTAGAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAA 4206

Qy        339 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 398
Db       4207 AGATCTTGAAGAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAAT 4266

Qy        399 AGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTT 458
Db       4267 TTTAGAAGATTATAAAGAATTAAAAACTATTGAAACAGATA---TTTTAGAAGAGAAAAA 4323

Qy        459 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 518
Db       4324 AGAAATAGAAAAAGATCATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCT 4383

Qy        519 ACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGA 578
Db       4384 TGAAGCAGATATATTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGA 4443

Qy        579 TTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACA 638
Db       4444 AGAAGTACACGAATTA------AAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCA 4497

Qy        639 AAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGC 698
Db       4498 TATAAAAGGTTTGGAAGAAGATGATTTAGAAGAAGTAGATGATTTAAAAGGAAGTATATT 4557

Qy        699 AGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACG 758
Db       4558 AGACATGTTAAAGGGAGATATGGAATTAGGGGATATGGATAAGGAAAGTTTAGAAGATGT 4617

Qy        759 TGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGA 818
Db       4618 AACAACAAAACTTGGAGAAAGAGTTGAATCCTTAAAAGATGTTTTATCTAGTGCATTAGG 4677

Qy        819 TTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGA 878
Db       4678 CATGGATGAAGAACAAAT------GAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTGGA 4731

Qy        879 AGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACC 938
Db       4732 AGAAGTATTATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAGAA 4791

Qy        939 AGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATAT 998
Db       4792 AGTAAGGTTGATATTAAGGATAAGGAACCAAAAGATGAAATAGTAGAAGTTGAAATGAA 4851

Qy        999 TTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGA 1058
Db       4852 AGATGAAGATATAGAAGAAGATGTAGAAGAAGATATAGAAGAAGATATAGAAGAAGATAA 4911

Qy       1059 CGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGT 1118
Db       4912 AGTTGAAGATATAGATGAAGATATA---GATGAAGATATAGGTGAAGACAAAGATGAAGT 4968

Qy       1119 GCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTT 1178
Db       4969 TATAGATTTAATAGTCCAAAAAGAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAA 5028

Qy       1179 GATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGA 1238
Db       5029 ATTAGAAAAAAAAGTTGAAGAAGGTGTTAGTGGTCTTAAAAAACACGTAGACGAAGTAAT 5088

Qy       1239 ATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTT 1298
Db       5089 GAAATATGTTCAAAAAATTGATAAAGAAGTTGATAAAGAAGTATCTAAAGCTTTAGAATC 5148

Qy       1299 TAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCA 1358
Db       5149 AAAAAATGATGTTACTAATGTTTTAAAACAAAATCAAGATTTTTTTAGTAAAGTTAAAAA 5208

Qy       1359 GGTTAATAAGGAAAAGGAAAAATTCATAAAATCATT 1394
Db       5209 CTTCGTAAAAAAATATAAAGTATTTGCTGCACCATT 5244
```

```
RESULT 27
US-09-949-016-11814/c
; Sequence 11814, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11814
;   LENGTH: 29927
;   TYPE: DNA
```

```
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(29927)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11814

  Query Match             6.2%;  Score 91.6;  DB 3;  Length 29927;
  Best Local Similarity   47.0%;  Pred. No. 3.7e-07;
  Matches  324;  Conservative    0;  Mismatches  359;  Indels    7;  Gaps    1;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              |||   | | || ||   |||| |||||| |       | ||| |    |  |||   | |
Db       4628 AGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGGAGAGAGACAAAGAGAGACAGAGA 4569

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              ||  |   |||   ||| |||        |   |          |||      | |  ||  |
Db       4568 CAGACAGAGAGACAGAGACAGAAAAAGAGAGAGAGGAGAGAGAGAAAGAGAGACAGAGAT 4509

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||   | ||||||     |||  | ||  |  | |   | | || | |  |||   |||
Db       4508 AGAGATACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAGAGGTAGAGACACAGA 4449

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              ||||       || || ||    |  | |  | || ||   |||   | |||||    |
Db       4448 GAGAGACAGAGAAAAAGAGACAGAGAGAGATGAGAGAGAGAAAGAGAGACAGAGATAGAC 4389

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
              |   | ||   ||   | ||  |       |||  |||| ||||    | |   | |
Db       4388 AGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAGACAGAGTCAGAGATA 4329

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
              | |   |    |||| ||   |||   ||||| ||       | || |    |  |||
Db       4328 CAGAGAGAGGAGAGAAAGAGAGACAGAGACAGAGATACAGAGAGAGGAGAGAAAGAGAGA 4269

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | | ||  |   | |   | ||||||      |||  | ||  |  | |   | |  ||
Db       4268 CAGAGACAGACAGAGATACAGACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAG 4209

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
              | |  |||   |||||||       || || ||    |  | |  | || ||   |||
Db       4208 AGGTAGAGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGAGAGGAGAGAGAGAAAGA 4149

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              | |||||    |  |   | ||   ||   | ||  |         |||  |||||||
Db       4148 GAGACAGAGATAGACAGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAG 4089

Qy        543 AC-------GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 595
              ||       | | || |||  ||  |  ||||  | | || ||   |||   |||||||
Db       4088 ACAGAGACAGAGATACAGAGTAGAGGAGAGAAAGAGACAGAGACAGAGATACAGAGAGAG 4029

Qy        596 GTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
              | |  | ||| |    |  |||     | |||||   |||  | || |||   || |
Db       4028 GAGAGAGAGAGACAGAGACAGACAGAGACAGAGAAAGAGAGGAGAGAAAGAGAGACAGAG 3969

Qy        656 AAAAAAATTTAGAAAGAAAAAAGGAACATG 685
              | | |   ||| ||| | |  ||| | |
Db       3968 GTAGAGACACAGAGAGAGACAGAGAAAAAG 3939


RESULT 28
US-09-949-016-17474/c
; Sequence 17474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17474
;   LENGTH: 29927
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(29927)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17474

  Query Match             6.2%;  Score 91.6;  DB 3;  Length 29927;
  Best Local Similarity   47.0%;  Pred. No. 3.7e-07;
  Matches  324;  Conservative    0;  Mismatches  359;  Indels    7;  Gaps    1;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              |||   | | || ||   |||| |||||| |       | ||| |    |  |||   | |
Db       4628 AGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGGAGAGAGACAAAGAGAGACAGAGA 4569

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              ||  |   |||   ||| |||        |   |          |||      | |  ||  |
Db       4568 CAGACAGAGAGACAGAGACAGAAAAAGAGAGAGAGGAGAGAGAGAAAGAGAGACAGAGAT 4509

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||   | ||||||     |||  | ||  |  | |   | | || | |  |||   |||
Db       4508 AGAGATACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAGAGGTAGAGACACAGA 4449

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              ||||       || || ||    |  | |  | || ||   |||   | |||||    |
Db       4448 GAGAGACAGAGAAAAAGAGACAGAGAGAGATGAGAGAGAGAAAGAGAGACAGAGATAGAC 4389

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
              |   | ||   ||   | ||  |       |||  |||| ||||    | |   | |
Db       4388 AGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAGACAGAGTCAGAGATA 4329

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
              | |   |    |||| ||   |||   ||||| ||       | || |    |  |||
Db       4328 CAGAGAGAGGAGAGAAAGAGAGACAGAGACAGAGATACAGAGAGAGGAGAGAAAGAGAGA 4269

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | | ||  |   | |   | ||||||      |||  | ||  |  | |   | |  ||
Db       4268 CAGAGACAGACAGAGATACAGACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAG 4209

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
              | |  |||   |||||||       || || ||    |  | |  | || ||   |||
Db       4208 AGGTAGAGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGAGAGGAGAGAGAGAAAGA 4149

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              | |||||    |  |   | ||   ||   | ||  |         |||  |||||||
Db       4148 GAGACAGAGATAGACAGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAG 4089

Qy        543 AC-------GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 595
              ||       | | || |||  ||  |  ||||  | | || ||   |||   |||||||
Db       4088 ACAGAGACAGAGATACAGAGTAGAGGAGAGAAAGAGACAGAGACAGAGATACAGAGAGAG 4029

Qy        596 GTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
              | |  | ||| |    |  |||     | |||||   |||  | || |||   || |
Db       4028 GAGAGAGAGAGACAGAGACAGACAGAGACAGAGAAAGAGAGGAGAGAAAGAGAGACAGAG 3969

Qy        656 AAAAAAATTTAGAAAGAAAAAAGGAACATG 685
              | | |   ||| ||| | |  ||| | |
Db       3968 GTAGAGACACAGAGAGAGACAGAGAAAAAG 3939


RESULT 29
US-09-949-016-17475/c
; Sequence 17475, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17475
;   LENGTH: 29927
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(29927)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17475

  Query Match             6.2%;  Score 91.6;  DB 3;  Length 29927;
  Best Local Similarity   47.0%;  Pred. No. 3.7e-07;
  Matches  324;  Conservative    0;  Mismatches  359;  Indels    7;  Gaps    1;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              |||   | | || ||   |||| |||||| |       | ||| |     |  |||   | |
Db       4628 AGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGGAGAGAGACAAAGAGAGACAGAGA 4569

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              ||  |   |||   ||| |||        |   |         |||     | |  ||  |
Db       4568 CAGACAGAGAGACAGAGACAGAAAAAGAGAGAGAGGAGAGAGAGAAAGAGAGACAGAGAT 4509

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||   | ||||||      |||  | ||  |  | |   | | || | |  |||  :|||
Db       4508 AGAGATACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAGAGGTAGAGACACAGA 4449

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              ||||       || || ||    |  | |  | || ||   |||   | |||||    |
Db       4448 GAGAGACAGAGAAAAAGAGACAGAGAGAGATGAGAGAGAGAAAGAGAGACAGAGATAGAC 4389

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
              |   | ||   ||   | ||  |         |||  |||| ||||    | |   | |
Db       4388 AGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAGACAGAGTCAGAGATA 4329

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
              | |   |    |||| ||   |||   ||||| ||      | || |    |  |||
Db       4328 CAGAGAGAGGAGAGAAAGAGAGACAGAGACAGAGATACAGAGAGAGGAGAGAAAGAGAGA 4269

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | | ||  |   | |   | ||||||      |||  | ||  |  | |   | | ||
Db       4268 CAGAGACAGACAGAGATACAGACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAG 4209

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
              | |  |||   |||||||          || || ||      | | |  | || ||   |||
Db       4208 AGGTAGAGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGAGGAGAGAGAGAAAGA 4149

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              |  |||||   |  |   | ||   ||   | ||  |         |||  |||||||
Db       4148 GAGACAGAGATAGACAGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAG 4089

Qy        543 AC-------GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 595
              ||       | | || |||  ||  |  ||||  | | || ||   |||   |||||||
Db       4088 ACAGAGACAGAGATACAGAGTAGAGGAGAGAAAGAGACAGAGACAGAGATACAGAGAGAG 4029

Qy        596 GTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
              | |  | |||  |   |  |||     | |||||   |||  | || |||   || |
Db       4028 GAGAGAGAGAGACAGAGACAGACAGAGACAGAGAAAGAGAGGAGAGAAAGAGAGACAGAG 3969

Qy        656 AAAAAAATTTAGAAAGAAAAAAGGAACATG 685
                | | |   |||  ||| | |  ||| | |
Db       3968 GTAGAGACACAGAGAGAGACAGAGAAAAAG 3939


RESULT 30
US-09-949-016-12639/c
; Sequence 12639, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12639
;   LENGTH: 37802
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(37802)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12639

  Query Match             6.2%;  Score 91.6;  DB 3;  Length 37802;
  Best Local Similarity   47.0%;  Pred. No. 3.7e-07;
  Matches  324;  Conservative    0;  Mismatches  359;  Indels    7;  Gaps    1;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              |||   | | || ||   |||| |||||| |       | ||| |     |  |||   | |
Db       4628 AGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGGAGAGAGACAAAGAGAGACAGAGA 4569

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              ||  |   |||   ||| |||        |   |         |||     | |  ||  |
Db       4568 CAGACAGAGAGACAGAGACAGAAAAAGAGAGAGAGGAGAGAGAGAAAGAGAGACAGAGAT 4509

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||   | ||||||      |||  | ||  |  | |   | | || | |  |||  :|||
Db       4508 AGAGATACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAGAGGTAGAGACACAGA 4449

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              ||||       || || ||    |  | |  | || ||   |||   | |||||    |
Db       4448 GAGAGACAGAGAAAAAGAGACAGAGAGAGATGAGAGAGAGAAAGAGAGACAGAGATAGAC 4389

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
              |   | ||   ||   | ||  |         |||  |||| ||||    | |   | |
Db       4388 AGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAGACAGAGTCAGAGATA 4329

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
              | |   |    |||| ||   |||   ||||| ||      | || |    |  |||
Db       4328 CAGAGAGAGGAGAGAAAGAGAGACAGAGACAGAGATACAGAGAGAGGAGAGAAAGAGAGA 4269

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | | ||  |   | |   | ||||||      |||  | ||  |  | |   | | ||
Db       4268 CAGAGACAGACAGAGATACAGACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAG 4209

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
              | |  |||   |||||||          || || ||      | | |  | || ||   |||
Db       4208 AGGTAGAGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGAGGAGAGAGAGAAAGA 4149

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              |  |||||   |  |   | ||   ||   | ||  |         |||  |||||||
Db       4148 GAGACAGAGATAGACAGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAG 4089
```

```
Qy        543 AC-------GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 595
              ||       | | || |||  ||  |  ||||  | | || ||   |||   |||||||
Db       4088 ACAGAGACAGAGATACAGAGTAGAGGAGAGAAAGAGACAGAGACAGAGATACAGAGAGAG 4029

Qy        596 GTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
              | |  | ||| |    |  |||     | |||||   |||  | || |||   || |
Db       4028 GAGAGAGAGAGACAGAGACAGACAGAGACAGAGAAAGAGAGGAGAGAAAGAGAGACAGAG 3969

Qy        656 AAAAAAATTTAGAAAGAAAAAAGGAACATG 685
                | | |   ||| ||| | |  ||| | |
Db       3968 GTAGAGACACAGAGAGAGACAGAGAAAAAG 3939
```

Search completed: December 20, 2005, 15:19:07
Job time : 344.317 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 10:43:25 ; Search time 1605.5 Seconds
(without alignments)
7633.272 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1482	100.0	1482	3	US-09-837-344-41	Sequence 41, Appl
	2	1482	100.0	1482	3	US-09-900-963-41	Sequence 41, Appl
	3	1480.4	99.9	1482	3	US-09-837-344-45	Sequence 45, Appl
	4	1480.4	99.9	1482	3	US-09-900-963-45	Sequence 45, Appl
	5	1446	97.6	1493	3	US-09-837-344-38	Sequence 38, Appl
	6	1446	97.6	1493	3	US-09-900-963-38	Sequence 38, Appl
	7	950	64.1	950	3	US-09-837-344-32	Sequence 32, Appl
	8	950	64.1	950	3	US-09-900-963-32	Sequence 32, Appl
	9	472.8	31.9	988	3	US-09-837-344-34	Sequence 34, Appl
	10	472.8	31.9	988	3	US-09-900-963-34	Sequence 34, Appl
	11	470.8	31.8	954	3	US-09-837-344-37	Sequence 37, Appl
	12	470.8	31.8	954	3	US-09-900-963-37	Sequence 37, Appl
	13	233.4	15.7	1300	7	US-10-668-749A-1	Sequence 1, Appli
c	14	153.2	10.3	1297	8	US-10-425-115-67687	Sequence 67687, A
c	15	151.2	10.2	6292	7	US-10-221-714A-461	Sequence 461, App
c	16	149.2	10.1	6767	7	US-10-221-714A-330	Sequence 330, App
c	17	146	9.9	3662	10	US-11-097-143-5152	Sequence 5152, Ap
c	18	145.8	9.8	1243	8	US-10-425-115-172717	Sequence 172717,
c	19	145	9.8	1204	7	US-10-437-963-77858	Sequence 77858, A
c	20	142.4	9.6	1081	8	US-10-425-115-16756	Sequence 16756, A
c	21	139.4	9.4	9539	5	US-10-239-676-52	Sequence 52, Appl
c	22	139.4	9.4	9539	6	US-10-240-453-54	Sequence 54, Appl
c	23	137.8	9.3	6668	6	US-10-311-455-1670	Sequence 1670, Ap
c	24	132.8	9.0	778	8	US-10-363-345A-2179	Sequence 2179, Ap
	25	132.8	9.0	778	8	US-10-363-345A-2180	Sequence 2180, Ap
c	26	132.8	9.0	778	9	US-10-363-483A-2179	Sequence 2179, Ap
	27	132.8	9.0	778	9	US-10-363-483A-2180	Sequence 2180, Ap
c	28	131.4	8.9	7442	7	US-10-221-714A-409	Sequence 409, App
c	29	129.8	8.8	921	8	US-10-425-115-38710	Sequence 38710, A
c	30	129.8	8.8	1762	8	US-10-363-345A-7529	Sequence 7529, Ap
	31	129.8	8.8	1762	8	US-10-363-345A-7530	Sequence 7530, Ap
c	32	129.8	8.8	1762	9	US-10-363-483A-7529	Sequence 7529, Ap
	33	129.8	8.8	1762	9	US-10-363-483A-7530	Sequence 7530, Ap
c	34	129.2	8.7	1995	8	US-10-363-345A-7533	Sequence 7533, Ap
	35	129.2	8.7	1995	8	US-10-363-345A-7534	Sequence 7534, Ap
c	36	129.2	8.7	1995	9	US-10-363-483A-7533	Sequence 7533, Ap
	37	129.2	8.7	1995	9	US-10-363-483A-7534	Sequence 7534, Ap
c	38	129	8.7	1200	8	US-10-363-345A-25801	Sequence 25801, A
	39	129	8.7	1200	8	US-10-363-345A-25802	Sequence 25802, A
c	40	129	8.7	1200	9	US-10-363-483A-25801	Sequence 25801, A
	41	129	8.7	1200	9	US-10-363-483A-25802	Sequence 25802, A
c	42	129	8.7	7597	6	US-10-311-455-986	Sequence 986, App
c	43	129	8.7	14006	6	US-10-311-455-1931	Sequence 1931, Ap
c	44	128.6	8.7	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	45	128.6	8.7	3673778	6	US-10-312-841-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-837-344-41
; Sequence 41, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 41:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1482 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
```

```
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..1482
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-837-344-41

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-196;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              ||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 2
US-09-900-963-41
; Sequence 41, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
```

```
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-045
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 41:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1482 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..1482
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-900-963-41

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-196;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              ||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 3
US-09-837-344-45
; Sequence 45, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;         APPLICANT: GUERIN-MARCHAND, Claudine
;                    DRUILHE, Pierre
;         TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                             HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                             OF STIMULATING THE T LYMPHOCYTES
;         NUMBER OF SEQUENCES: 46
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Burns, Doane, Swecker & Mathis
;              STREET: P.O. Box 1404
;              CITY: Alexandria
;              STATE: Virginia
;              COUNTRY: United States
;              ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
```

```
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/837,344
;              FILING DATE: 19-Apr-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/462,625
;              FILING DATE: <Unknown>
;              APPLICATION NUMBER: FR 91 01286
;              FILING DATE: 05-FEB-1991
;         ATTORNEY/AGENT INFORMATION:
;              NAME: McGowan, Malcolm K.
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-078
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 45:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1482 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..1482
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-837-344-45

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 4e-196;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 4
US-09-900-963-45
; Sequence 45, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
```

```
;                 DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 45:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1482 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1482
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-900-963-45

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 4e-196;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
```

```
Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 5
US-09-837-344-38
; Sequence 38, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1493 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-837-344-38

  Query Match             97.6%;  Score 1446;  DB 3;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-191;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257
```

```
Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 6
US-09-900-963-38
; Sequence 38, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 38:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1493 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-900-963-38

  Query Match             97.6%;  Score 1446;  DB 3;  Length 1493;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-191;
  Matches 1479;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137
```

```
Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479
```

```
RESULT 7
US-09-837-344-32
; Sequence 32, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-837-344-32

  Query Match             64.1%;  Score 950;  DB 3;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-122;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950
```

```
RESULT 8
US-09-900-963-32.
; Sequence 32, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-900-963-32

  Query Match             64.1%;  Score 950;  DB 3;  Length 950;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-122;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy         611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy         671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy         731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy         791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy         851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy         911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 9
US-09-837-344-34
; Sequence 34, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
```

```
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 34:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 988 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-837-344-34

  Query Match             31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity   95.7%;  Pred. No. 1.7e-56;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db         661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
               |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db         841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy         497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy         557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
               ||||||||||||||||||||||||||||
Db         961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 10
US-09-900-963-34
; Sequence 34, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 34:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 988 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-900-963-34

  Query Match             31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity   95.7%;  Pred. No. 1.7e-56;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db         661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
```

```
                    ||||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 11
US-09-837-344-37
; Sequence 37, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 37:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 954 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..954
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-837-344-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 3.2e-56;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||

Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db        509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||| |||||||||||||||||| ||||| || |
Db        629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db        809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

Qy        557 AGTTGCAAGAACAACAAAGCGATTTA 582
              |||||||||||||||||||||||||
Db        929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 12
US-09-900-963-37
; Sequence 37, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30.
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 37:
```

```
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 954 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..954
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-900-963-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 3.2e-56;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db         629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
               |||||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db         809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy         497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

Qy         557 AGTTGCAAGAACAACAAAGCGATTTA 582
               |||||||||||||||||||||||||
Db         929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 13
US-10-668-749A-1
; Sequence 1, Application US/10668749A
; Publication No. US20040110205A1
; GENERAL INFORMATION:
;  APPLICANT: Agilent Technologies
;  TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis
;  FILE REFERENCE: 50112-1580
;  CURRENT APPLICATION NUMBER: US/10/668,749A
;  CURRENT FILING DATE:  2003-09-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic construct
US-10-668-749A-1

  Query Match             15.7%;  Score 233.4;  DB 7;  Length 1300;
  Best Local Similarity   48.7%;  Pred. No. 2.8e-23;
  Matches  633;  Conservative    0;  Mismatches  666;  Indels    0;  Gaps    0;

Qy          91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 150
               ||| ||||  | || |  |  |||   |   | || || | | |      ||| ||||
Db           1 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy         151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
               || || || |||   |   | || || | | |      ||| ||||     || ||
Db          61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy         211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
               || |||   |   | || || | | |      ||| ||||     || || || |||
Db         121 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 180

Qy         271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
               |   | || || | | |      ||| ||||  | || |  | |||   |   | ||
Db         181 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240

Qy         331 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
               || | | |      ||| ||||     || || || |||   |   | || || | | |
Db         241 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 300

Qy         391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAA 450
               ||| ||||     || || || |||   |   | || || | | |      |||
Db         301 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 360

Qy         451 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
               ||||     || || || |||   |   | || || | | |      ||| ||||
Db         361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 420

Qy         511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
               || || || |||   |   | || || | | |      ||| ||||     || || ||
Db         421 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 480

Qy         571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
               |||   |   | || || | | |      ||| ||||     || |  |  ||| | |
Db         481 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 540

Qy         631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
               | || |||  ||     | |  |||||||   | ||| ||||||  || |   | |
Db         541 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 600

Qy         691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
               | |  |  | |  |   | |          | ||| |  |   | | ||     | | ||
Db         601 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 660

Qy         751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
               ||  ||      | |  | | |  ||        |     | | || | |   ||
Db         661 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 720

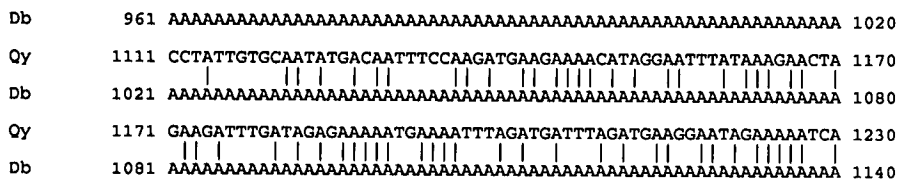
Qy         811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
               |  | | |   ||  ||| |   | || | |||||| ||| | | ||   |  | || |
Db         721 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 780

Qy         871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930
               ||      ||  |  ||   | | | | |||  | |      || |||| ||| |      |
Db         781 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840

Qy         931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
               | ||||  | || ||||| || | |||       | | | ||||  | | |  | | | |
Db         841 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 900

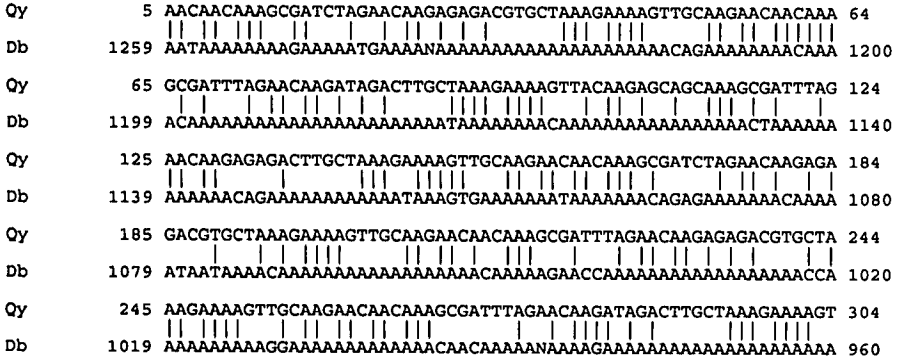
Qy         991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCT 1050
               |||  |      |   ||  |   ||| ||  ||  |  |  |  ||| ||
Db         901 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 960

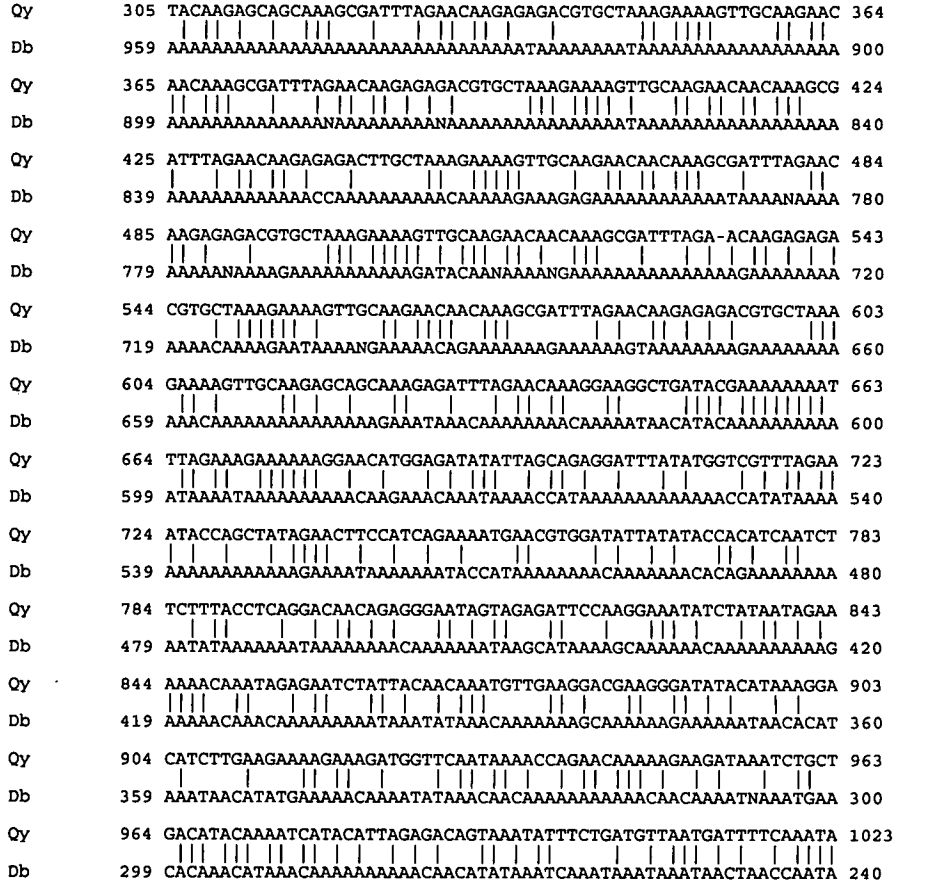
Qy        1051 GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAG 1110
                || |  |  |   |  || | |  || || || |  |  || |   | |  ||   ||
```



RESULT 14
US-10-425-115-67687/c
; Sequence 67687, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67687
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1297)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_161728C.1
US-10-425-115-67687

Query Match 10.3%; Score 153.2; DB 8; Length 1297;
Best Local Similarity 47.2%; Pred. No. 3.6e-12;
Matches 481; Conservative 0; Mismatches 538; Indels 1; Gaps 1;

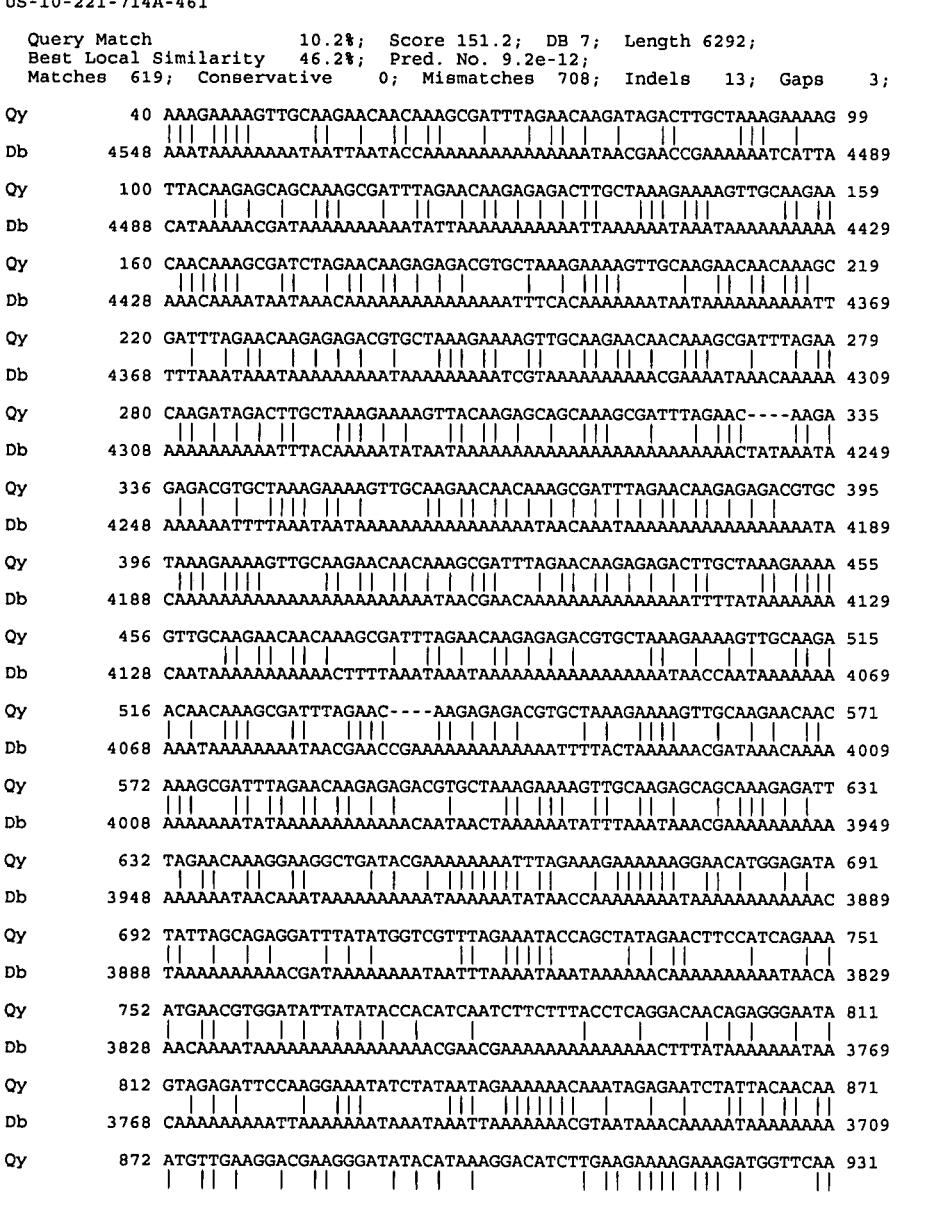




RESULT 15
US-10-221-714A-461/c
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30

```
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
;   LENGTH: 6292
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

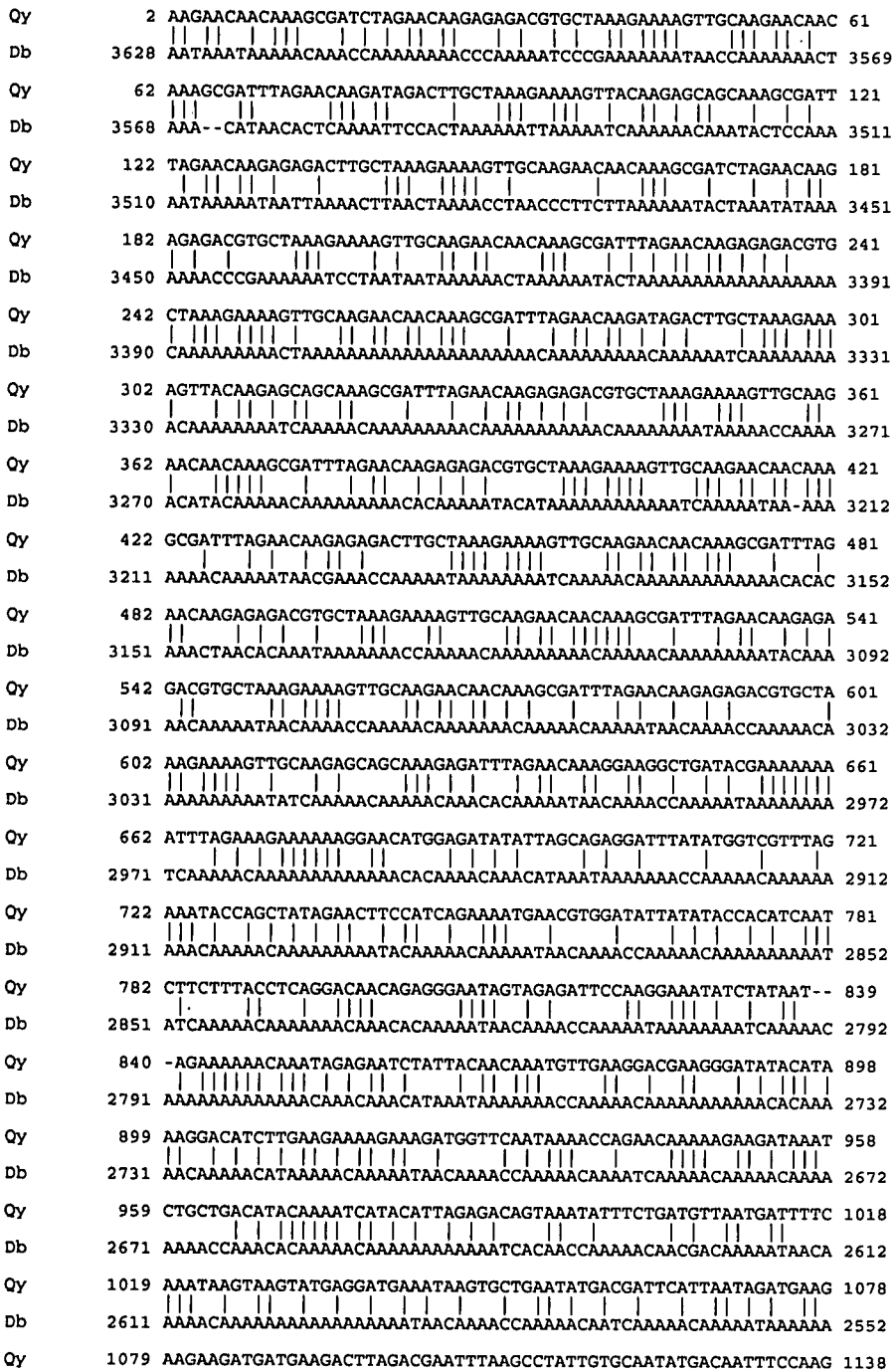
  Query Match            10.2%;  Score 151.2;  DB 7;  Length 6292;
  Best Local Similarity   46.2%;  Pred. No. 9.2e-12;
  Matches  619;  Conservative    0;  Mismatches  708;  Indels   13;  Gaps    3;
```





```
RESULT 16
US-10-221-714A-330/c
; Sequence 330, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 330
;   LENGTH: 6767
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-330

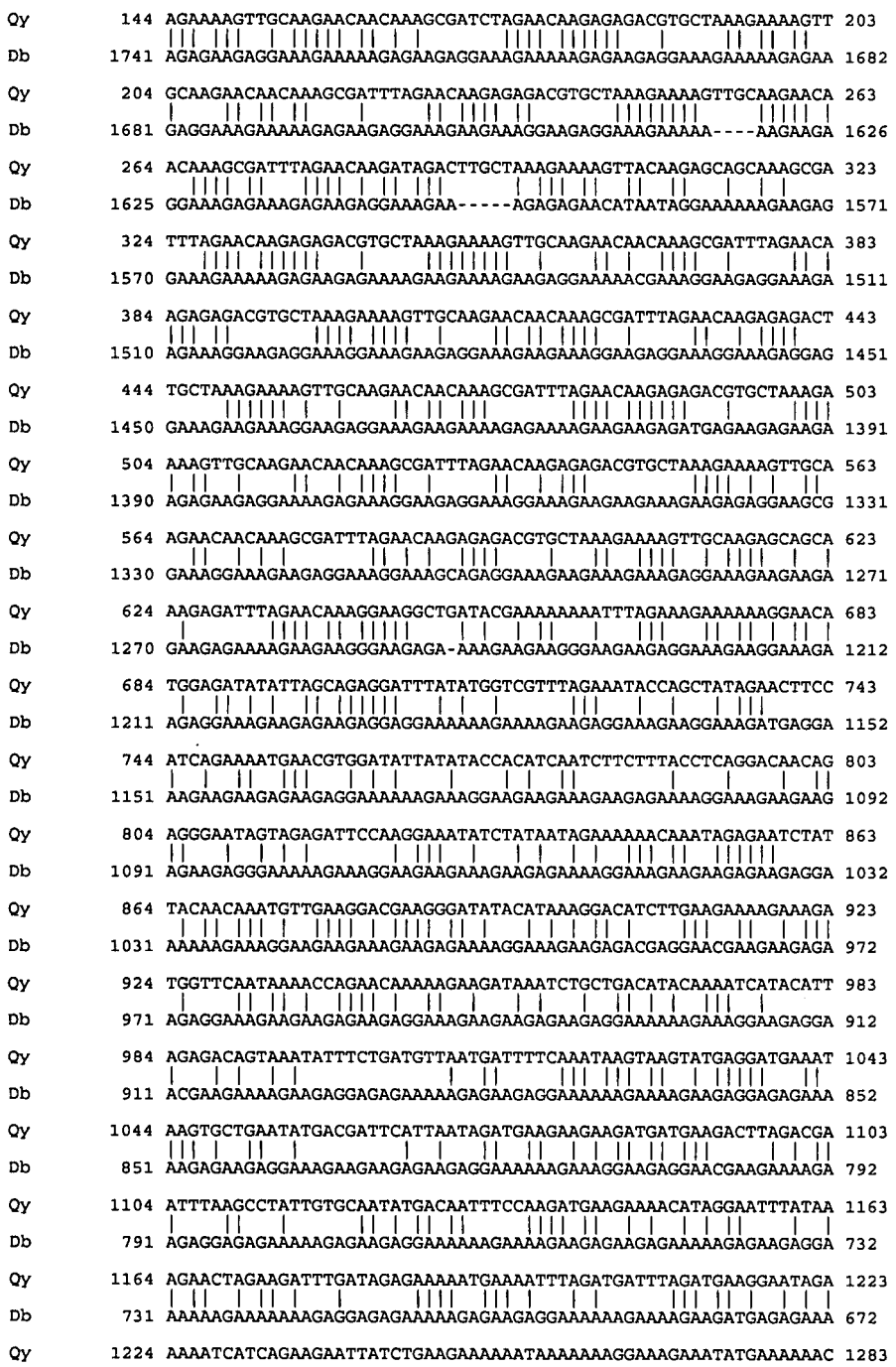
  Query Match            10.1%;  Score 149.2;  DB 7;  Length 6767;
  Best Local Similarity   46.1%;  Pred. No. 1.8e-11;
  Matches  643;  Conservative    0;  Mismatches  743;  Indels    8;  Gaps    4;
```

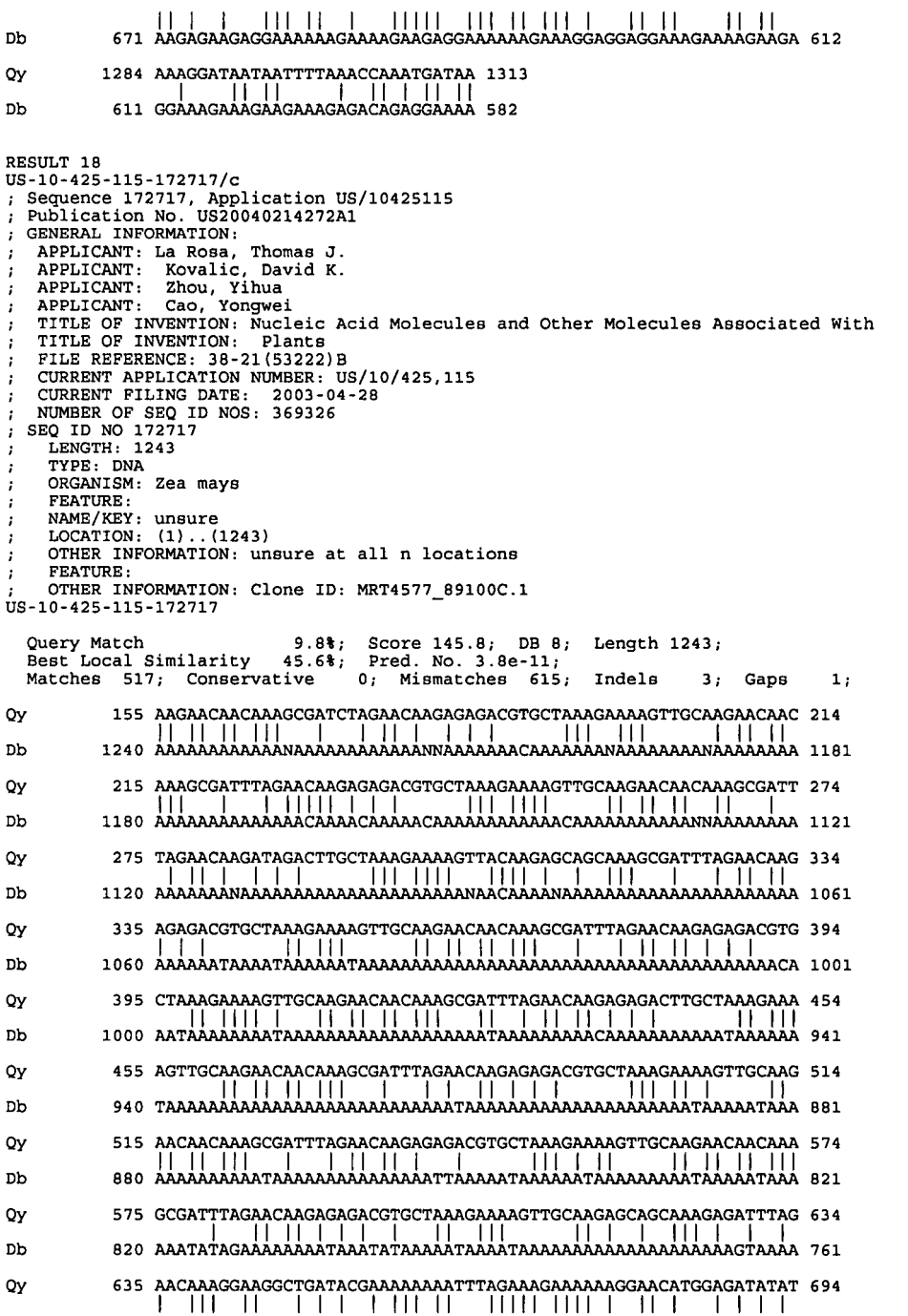


RESULT 17
US-11-097-143-5152/c
; Sequence 5152, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5152
; LENGTH: 3662
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-5152

Query Match 9.9%; Score 146; DB 10; Length 3662;
Best Local Similarity 46.7%; Pred. No. 4.4e-11;
Matches 603; Conservative 0; Mismatches 675; Indels 12; Gaps 4;

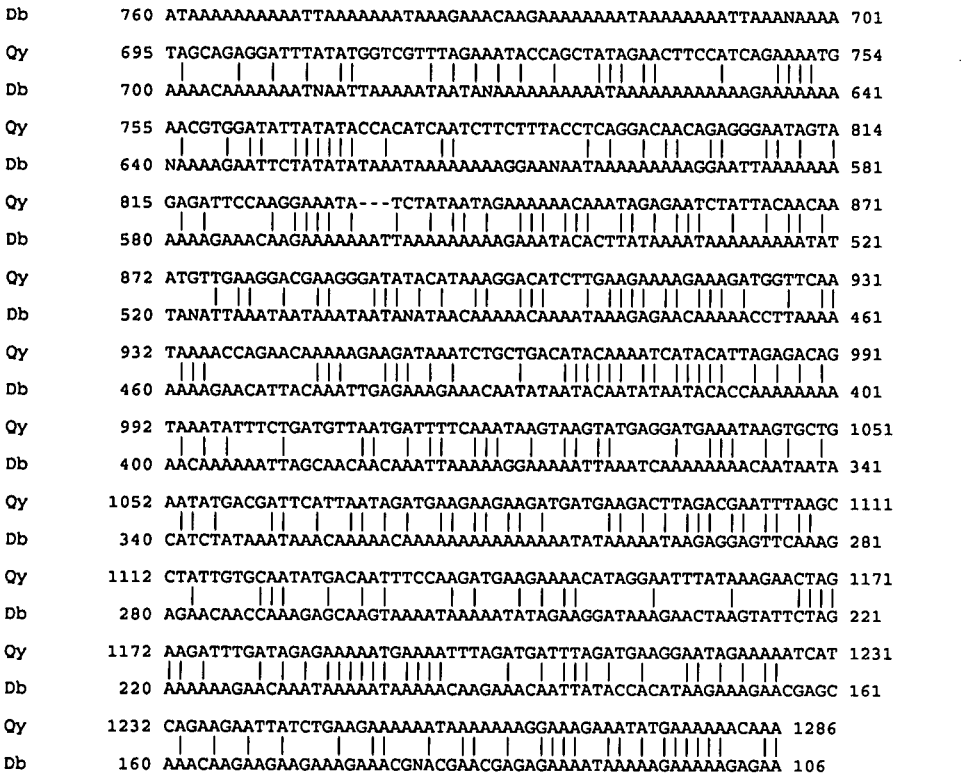




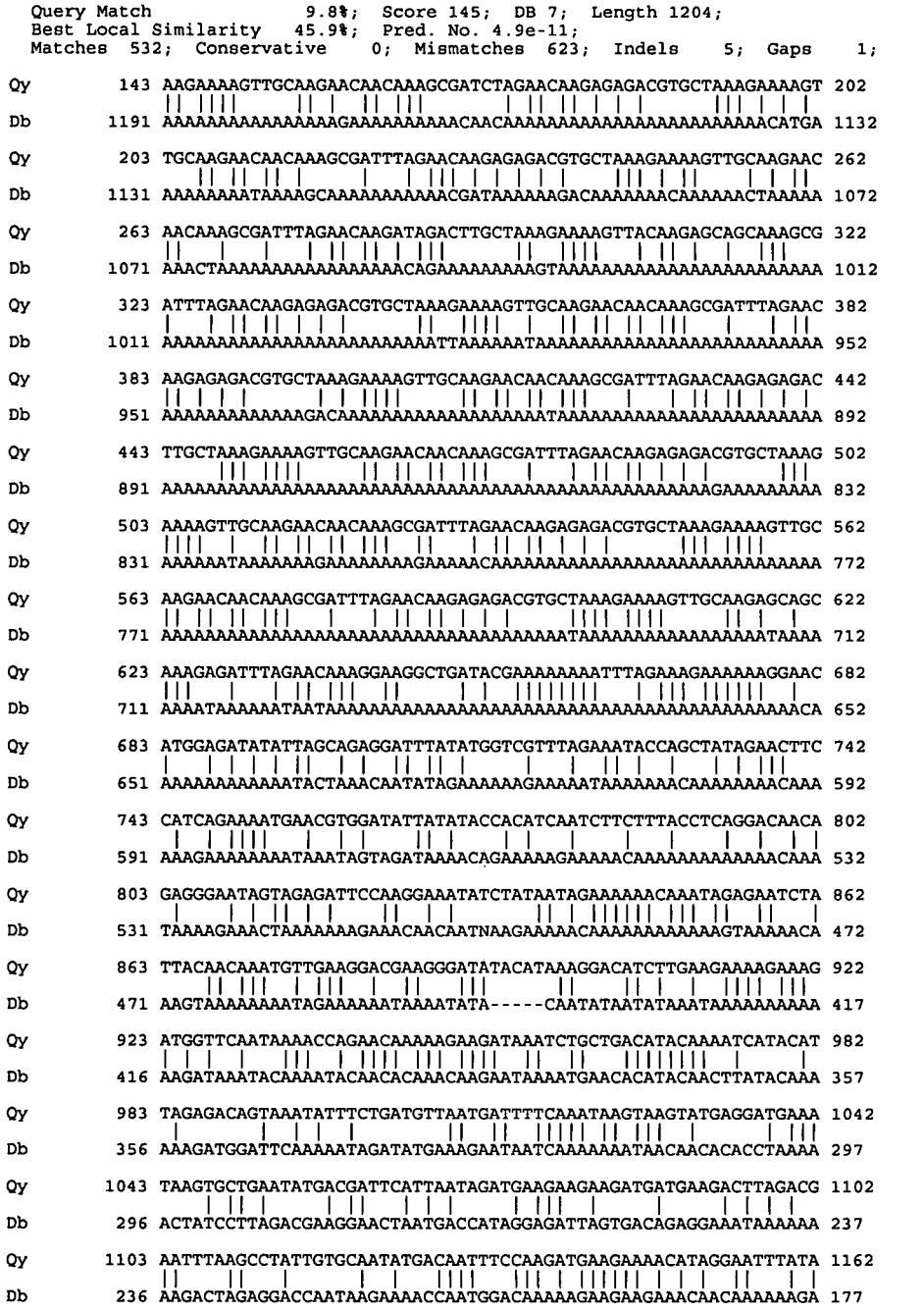


RESULT 18
US-10-425-115-172717/c
; Sequence 172717, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172717
; LENGTH: 1243
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1243)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717

Query Match 9.8%; Score 145.8; DB 8; Length 1243;
Best Local Similarity 45.6%; Pred. No. 3.8e-11;
Matches 517; Conservative 0; Mismatches 615; Indels 3; Gaps 1;



RESULT 19
US-10-437-963-77858/c
; Sequence 77858, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 77858
; LENGTH: 1204
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1204)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_77718C.1
US-10-437-963-77858



```
Qy       1163 AAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAG 1222
Db        176 AAATAGAAGAAAAACACAGAGAAGAAAAAAANCGAGAAAACAACGAACNCAAAAGAACAA 117

Qy       1223 AAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAA 1282
Db        116 TATAAGACGAGAAACGAACGAGGAAAGTACGAAACGACAGAAGAAATATAAACAAGAGAA 57

Qy       1283 CAAAGGATAATAATTTTAAA 1302
Db          56 CACCCGAGTAGAAGGTCGAA 37
```

RESULT 20
US-10-425-115-16756/c
; Sequence 16756, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 16756
; LENGTH: 1081
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1081)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_115279C.1
US-10-425-115-16756

Query Match 9.6%; Score 142.4; DB 8; Length 1081;
Best Local Similarity 47.3%; Pred. No. 1.1e-10;
Matches 413; Conservative 0; Mismatches 460; Indels 0; Gaps 0;

```
Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
Db       1067 AAAAAAAGGAATAAAGAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 1008

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
Db       1007 TAAAAGTAAAAGAAACAAAAAAATAAAAAAAAACACACAAAAGAAAAAAACAAATAAAAA 948

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db        947 AAAATAAAAGAAAAAAGAAAAAAAAAGAAAAAAAAAAAAGTAAAAAAAAAAAAAACAAAA 888

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db        887 AAAAAAACAAAAAAAGANAAAAAAACCAAAATAAACAAAAGAAAATAAAAACCAAAAAAA 828

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
Db        827 AAAAAAAAAAAAAAAAAAAAAAACAAAAAAGAAAAAAAAAAAAAAAAAAGCAAAAAAAAAC 768

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
Db        767 AAAAAAAAAAAAAAAATAAATAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAA 708

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
Db        707 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAACAAAAAAAAAGAAAAAA 648

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
Db        647 AAAAACAATAAAATAAAATAATATAAAAAAAAAACANAAAAACAAACAAAAATAAAGAATA 588

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
Db        587 AAAAATAAAACAAAAAGNNAAAACAGNAAAAAAAACANATAAACAAAAAAAAAAATAGAAT 528

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
Db        527 AAAAAAGAAANAAAAAATACAANAAATAAGAAAAAACATAATAAACAAAAAAAAAACAAA 468

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
Db        467 AAAAATATGAAAAAATAACAAAAAAACACNAAAAAAAACAAAAAAAAAAAAAAAAACAA 408

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
Db        407 TATAATAAGAAAAAAAAAAAAAAAAAAAAAGATATAAAAAAATAAAAGAAGAAAAAAAAA 348

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
Db        347 AAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAACAAAAAAAAAAATAAAAA 288

Qy       1330 CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db        287 AAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAACAAAAAAACAGAAAAA 228

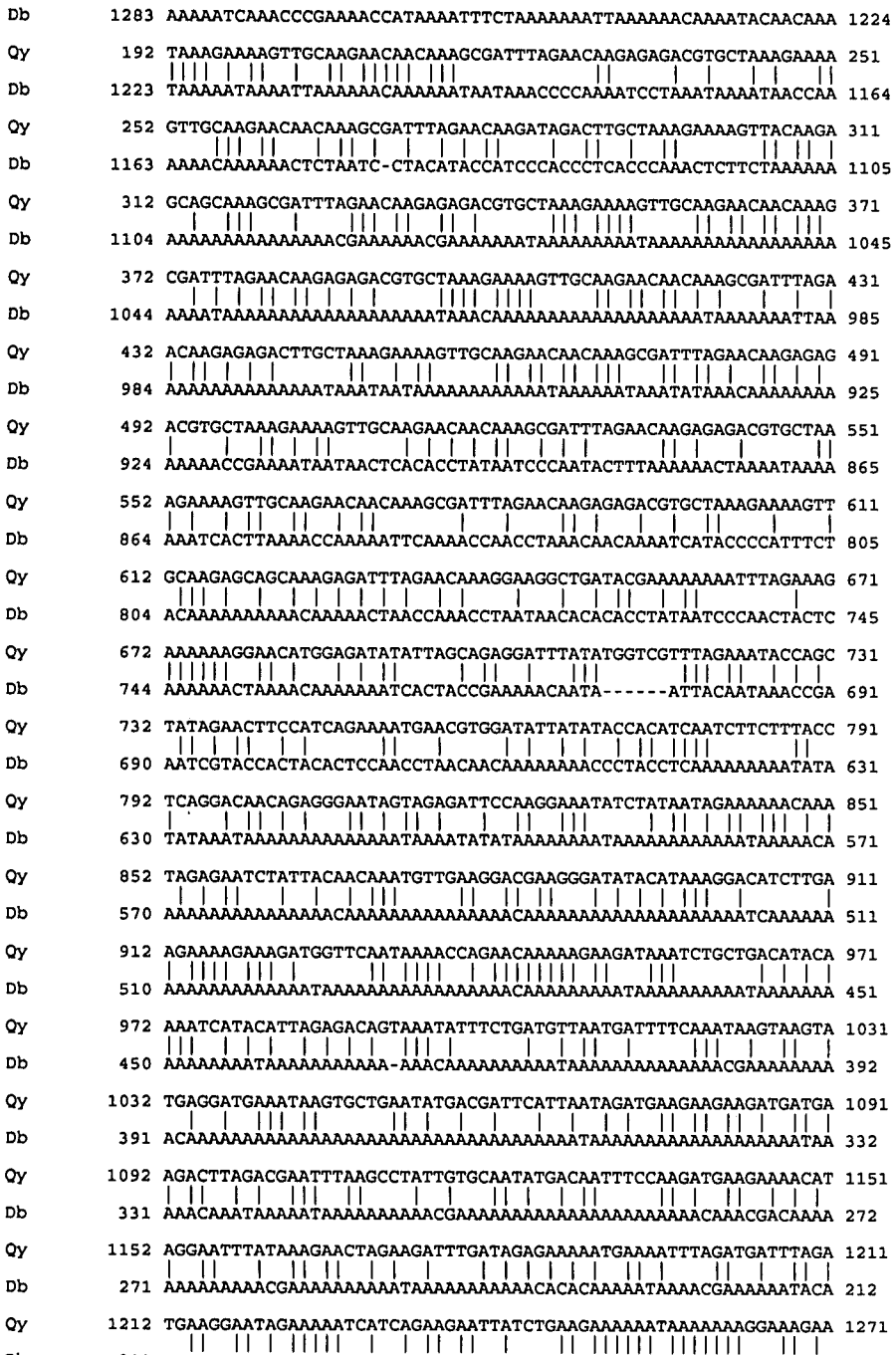
Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAA 1422
Db        227 GACTCATAACAAACATACAAAAAACAAAAAGCA 195

RESULT 21
US-10-239-676-52/c
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 52
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-52

Query Match 9.4%; Score 139.4; DB 5; Length 9539;
Best Local Similarity 45.6%; Pred. No. 4.3e-10;
Matches 607; Conservative 0; Mismatches 716; Indels 8; Gaps 3;

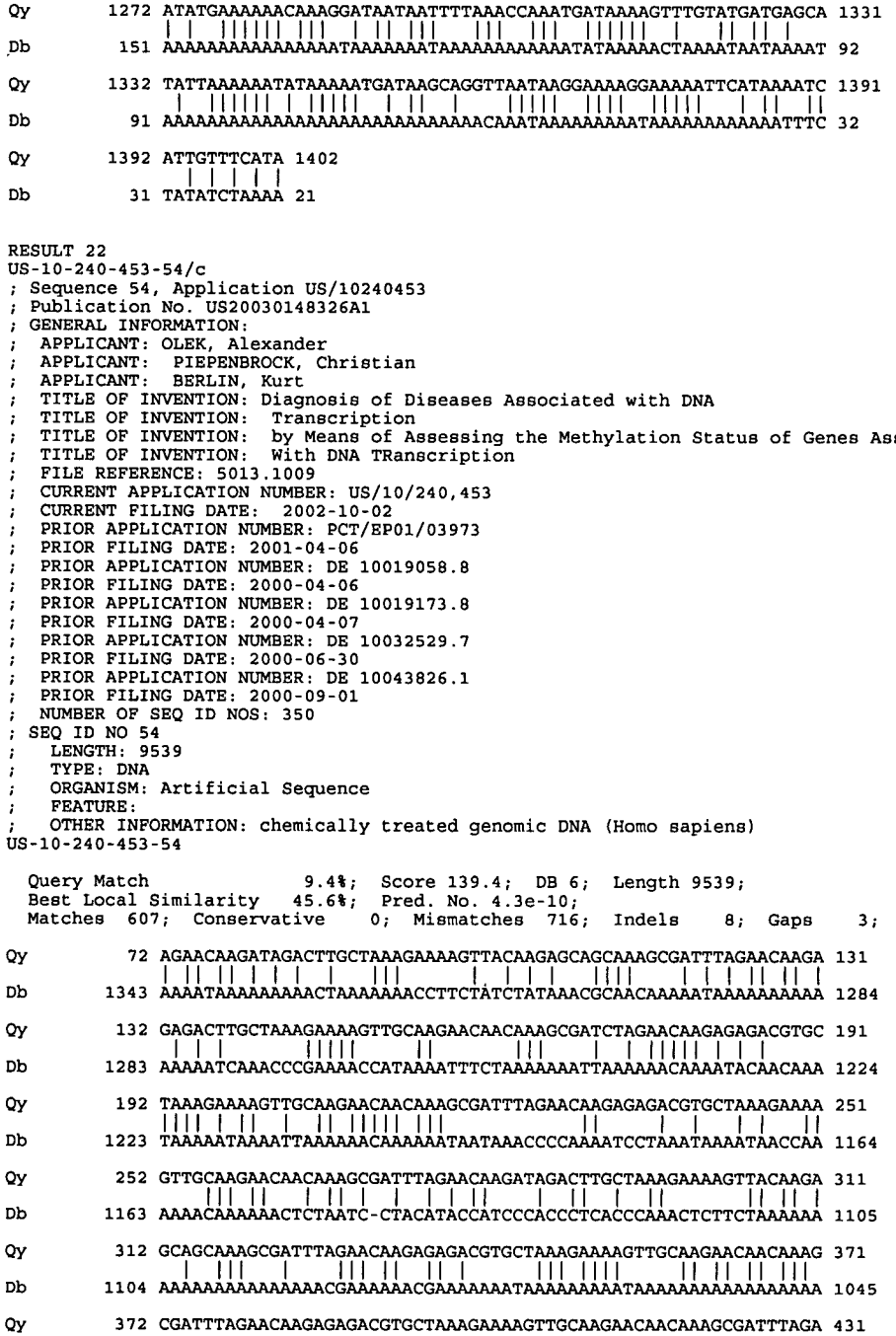
```
Qy   72 AGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGA 131
Db   1343 AAAATAAAAAAAAACTAAAAAAACCTTCTATCTATAAACGCAACAAAAATAAAAAAAAAA 1284

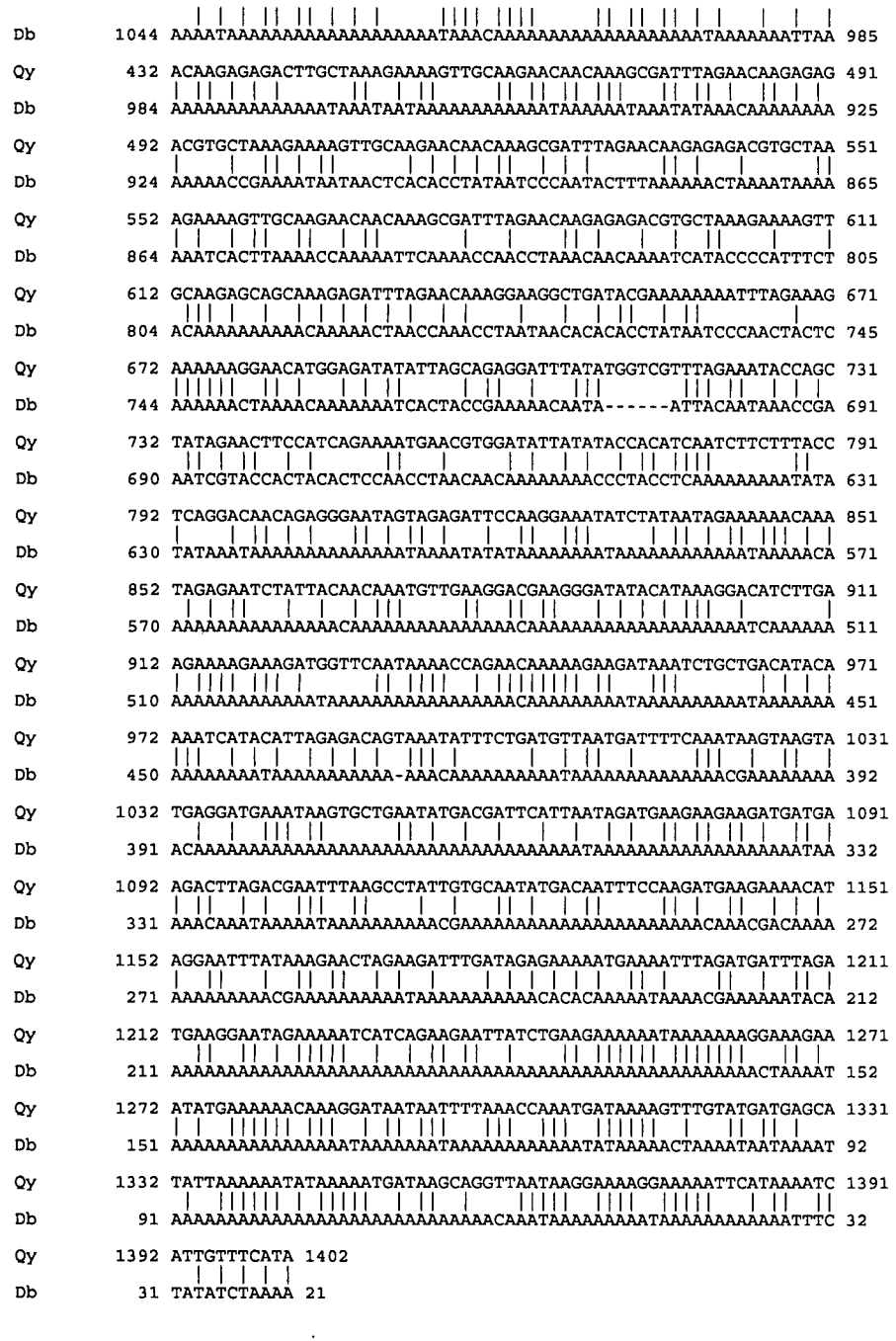
Qy   132 GAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGC 191
```
```



RESULT 22
US-10-240-453-54/c
; Sequence 54, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA TRanscription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 54
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-54

Query Match 9.4%; Score 139.4; DB 6; Length 9539;
Best Local Similarity 45.6%; Pred. No. 4.3e-10;
Matches 607; Conservative 0; Mismatches 716; Indels 8; Gaps 3;

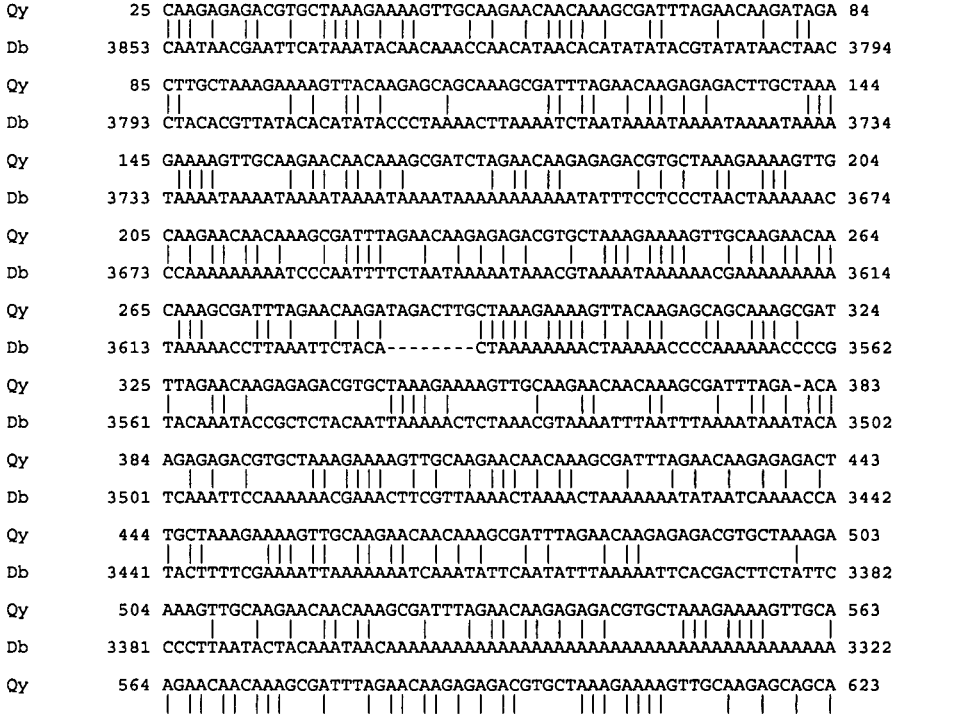


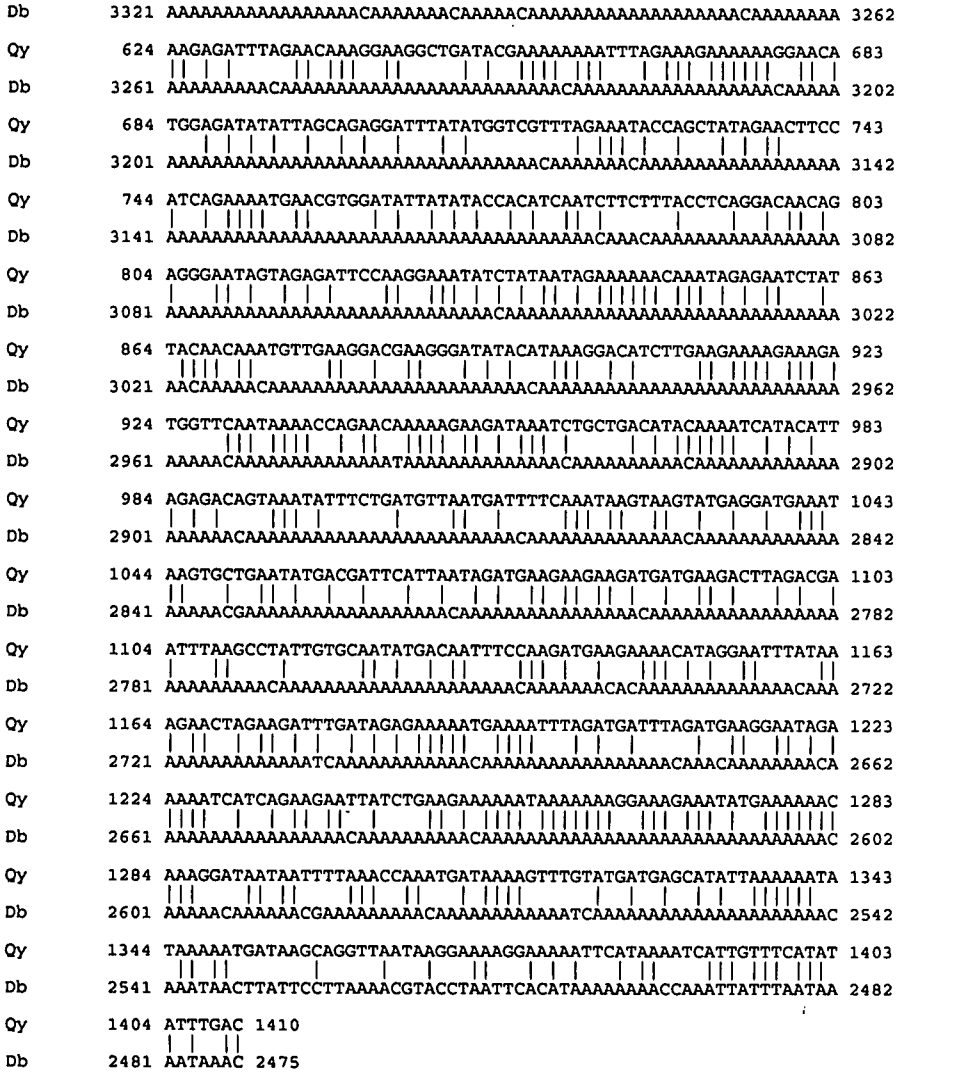


RESULT 23

US-10-311-455-1670/c
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1670
; LENGTH: 6668
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1936
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

Query Match 9.3%; Score 137.8; DB 6; Length 6668;
Best Local Similarity 44.8%; Pred. No. 6.7e-10;
Matches 621; Conservative 0; Mismatches 757; Indels 9; Gaps 2;





RESULT 24
US-10-363-345A-2179/c
; Sequence 2179, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712

; SEQ ID NO 2179
; LENGTH: 778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 2179
US-10-363-345A-2179

Query Match 9.0%; Score 132.8; DB 8; Length 778;
Best Local Similarity 48.5%; Pred. No. 2.2e-09;
Matches 365; Conservative 0; Mismatches 387; Indels 0; Gaps 0;

RESULT 25
US-10-363-345A-2180
; Sequence 2180, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock

```
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2180
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2180
US-10-363-345A-2180

  Query Match             9.0%;  Score 132.8;  DB 8;  Length 778;
  Best Local Similarity   48.5%;  Pred. No. 2.2e-09;
  Matches  365;  Conservative    0;  Mismatches  387;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||   | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db          6 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 65

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
              | |   | |       ||  |||| |  |   | | |      | | ||| |  |
Db         66 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 125

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
              | |  |  ||| | |  ||  |  | ||    | | | | |  || |  | |
Db        126 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAAATTATAAAAAAAAAAAAAA 185

Qy        818 ATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  ||| |   || ||| |||||| ||| |   |   |  | || ||||   |
Db        186 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTACG 245

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |   ||    ||||
Db        246 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAA 305

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | ||||  | |      | | |  ||| |
Db        306 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 365

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |   ||  ||| | ||| || ||| || ||  |  ||| ||       | |
Db        366 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 425

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   || || || || || || |  | |   |   ||    |
Db        426 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTACGAAAAAAAAAT 485

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||          || ||||  || |||   | | |  || | || |
Db        486 TACGAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAATATTACGAAAAAAA 545

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |  | |||  ||||   | || |   | |  ||  || | |  | |  | | ||
Db        546 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 605

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | || | |||  | || ||||
Db        606 AAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAAATT 665

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
              ||||  |||  | |  |      |  | |  ||  |||||||| | |||||     |
Db        666 ATAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTATAAA 725

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db        726 AAAAAAAAAATTATAAAAAAAAAAAAATTACAA 757
```

```
RESULT 26
US-10-363-483A-2179/c
; Sequence 2179, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2179
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2179
US-10-363-483A-2179

  Query Match             9.0%;  Score 132.8;  DB 9;  Length 778;
  Best Local Similarity   48.5%;  Pred. No. 2.2e-09;
  Matches  365;  Conservative    0;  Mismatches  387;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||   | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db        773 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 714

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
              | |   | |       ||  |||| |  |   | | |      | | ||| |  |
Db        713 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 654

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
              | |  |  ||| | |  ||  |  | ||    | | | | |  || |  | |
Db        653 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAATTATAAAAAAAAAAAAAA 594

Qy        818 ATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  ||| |   || ||| |||||| ||| |   |   |  | || ||||   |
Db        593 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACG 534

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |   ||    ||||
Db        533 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAA 474

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | ||||  | |      | | |  ||| |
Db        473 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 414

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |   ||  ||| | ||| || ||| || ||  |  ||| ||       | |
Db        413 TTACAAAAACGAAAAATTATAAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 354

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   || || || || || || |  | |   |   ||    |
Db        353 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTACGAAAAAAAAAT 294

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||          || ||||  || |||   | | |  || | || |
Db        293 TACGAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAA 234

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |  | |||  ||||   | || |   | |  ||  || | |  | |  | | ||
Db        233 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 174

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | || | |||  | || ||||
Db        173 AAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAAATT 114

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
              ||||  |||  | |  |      |  | |  ||  |||||||| | |||||     |
```

```
Db        113 ATAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAAATTATAAA 54

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db         53 AAAAAAAAAATTATAAAAAAAAAAAATTACAA 22


RESULT 27
US-10-363-483A-2180
; Sequence 2180, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2180
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2180
US-10-363-483A-2180

  Query Match             9.0%;  Score 132.8;  DB 9;  Length 778;
  Best Local Similarity   48.5%;  Pred. No. 2.2e-09;
  Matches  365;  Conservative    0;  Mismatches  387;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||   | |    |||||||||||   ||  ||||||  |  ||  | | | | |
Db          6 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 65

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
                | |   | |      ||  |||| |  |   | |  |      |  | ||| |  |
Db         66 TTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 125

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
                | |  | ||| | |  ||  |  | ||    | | |  | |   || | | |
Db        126 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAAATTATAAAAAAAAAAAAA 185

Qy        818 ATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  ||| |   || ||| |||||| ||| |    |    |  | || ||||   |
Db        186 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACG 245

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |   ||    ||||
Db        246 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTACGAAAAAA 305

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | |||| | |      | | |  ||| |
Db        306 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAAATTACGAAAAAAAAAAAAA 365

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |    ||  ||| | ||| || ||| || ||  |  ||| ||        | |
Db        366 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 425

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   || || || || || || | |  |   |   ||    |
Db        426 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTACGAAAAAAAAAT 485

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||           || |||| || |||   | | |  || | || | |
Db        486 TACGAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAAA 545

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |  | |||  ||||   | || |  | |  ||  || | |  |  |  | | ||
Db        546 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 605

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | || | |||  | || ||||
Db        606 AAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAATT 665

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
                ||||  |||  | |  |       |  |  |  ||  ||||||| | |||||     |
Db        666 ATAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTATAAA 725

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db        726 AAAAAAAAAATTATAAAAAAAAAAAAATTACAA 757


RESULT 28
US-10-221-714A-409/c
; Sequence 409, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 409
;   LENGTH: 7442
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1792, 4861..4862
;   OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-409

  Query Match             8.9%;  Score 131.4;  DB 7;  Length 7442;
  Best Local Similarity   46.7%;  Pred. No. 5.3e-09;
  Matches  565;  Conservative    0;  Mismatches  626;  Indels   18;  Gaps    4;

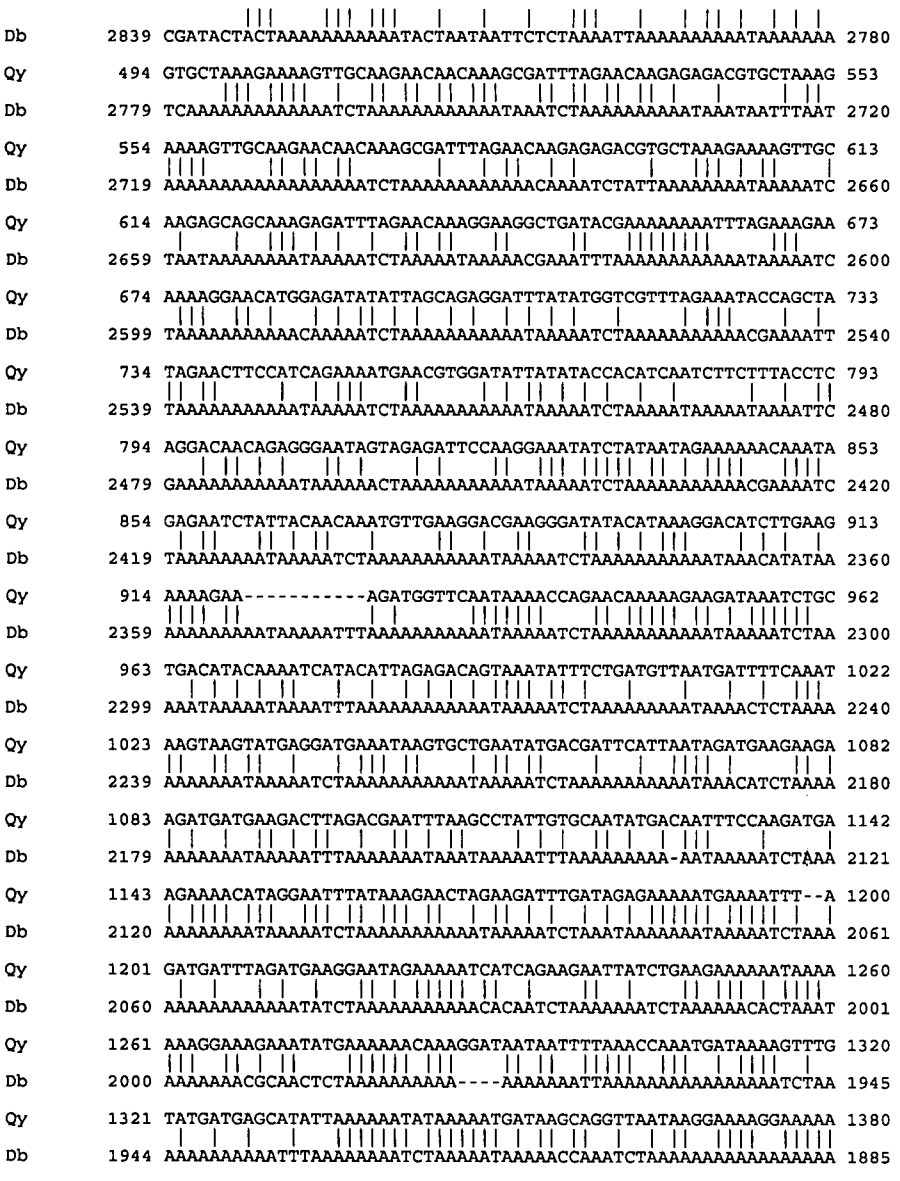
Qy        194 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 253
              || ||   |   |  | || |||   ||  |  || |      |    | ||| ||||
Db       3079 AAAAACCCTCAAACACAAAAAAAAAATATAAAACACTATCCCAATACACAAAAAAAAAAA 3020

Qy        254 TGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGC 313
                ||   || ||| |||     || |  | |  | |  ||| ||||   | || |
Db       3019 CTAAAATTTAAAAAAACGAAACTTAAAATAAAAACCCTCAAAAATAAAATAAAAAAAACA 2960

Qy        314 AGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 373
              |||| | ||       |  | | |      ||  ||      ||   |||| | |
Db       2959 CCCAAACCTATCCCACTCCTAACATCCCAAAACAACTAATACAATAATTCCAACCTAACA 2900

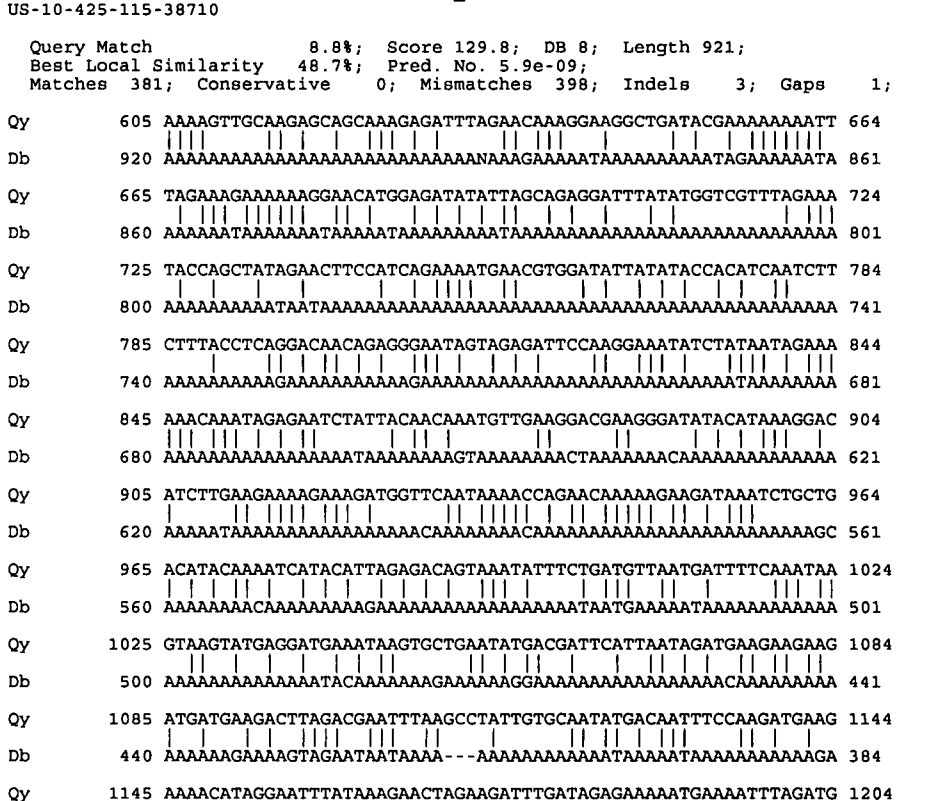
Qy        374 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 433
              |      | || | | |     |||  ||      || |     |||       | ||
Db       2899 AAACTATAAAAAAAAAAAAATATAATAAATCCGATAAAAACTTCGAAAAACTAAAAAAAA 2840

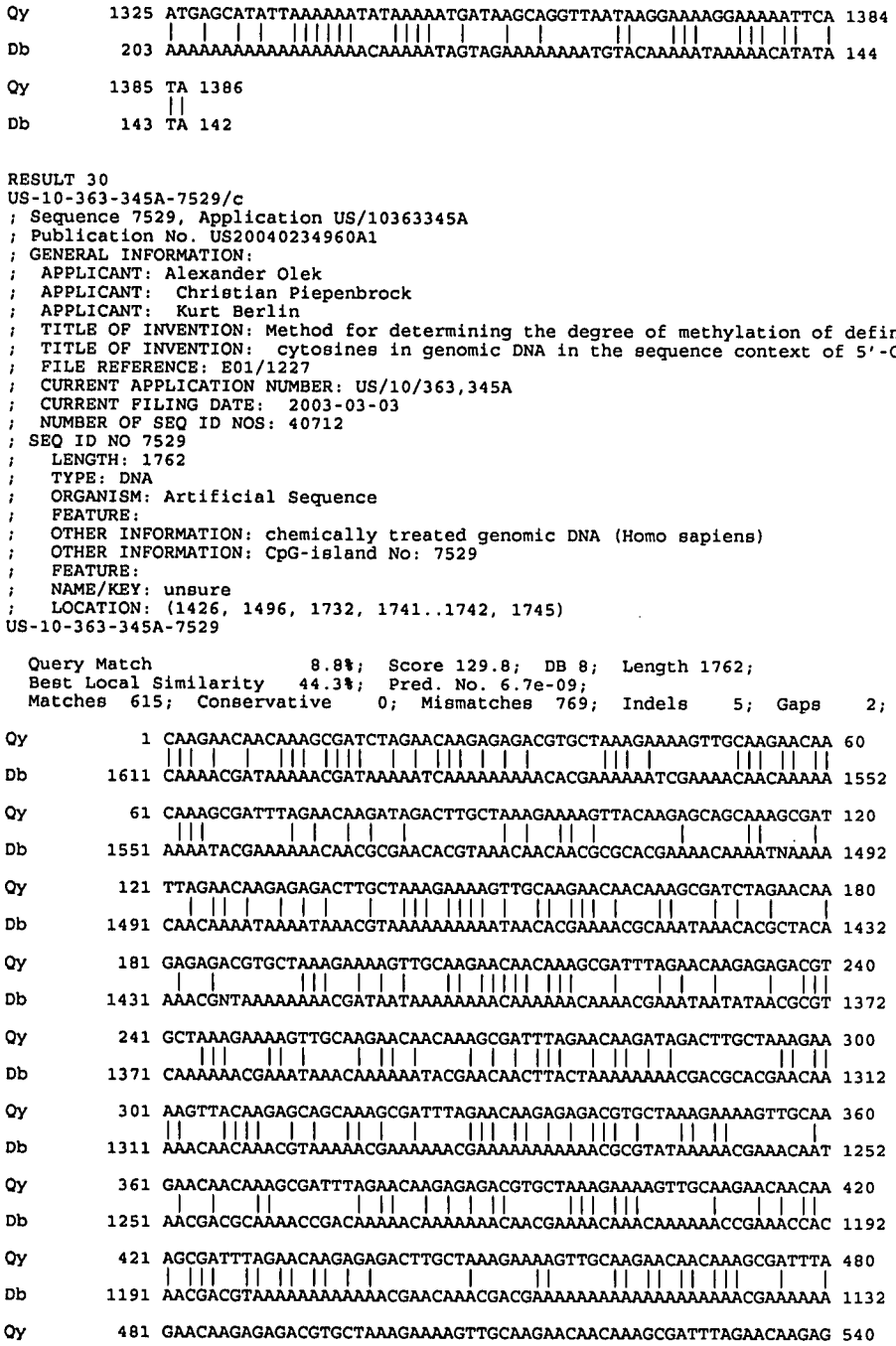
Qy        434 AAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 493
```



RESULT 29
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

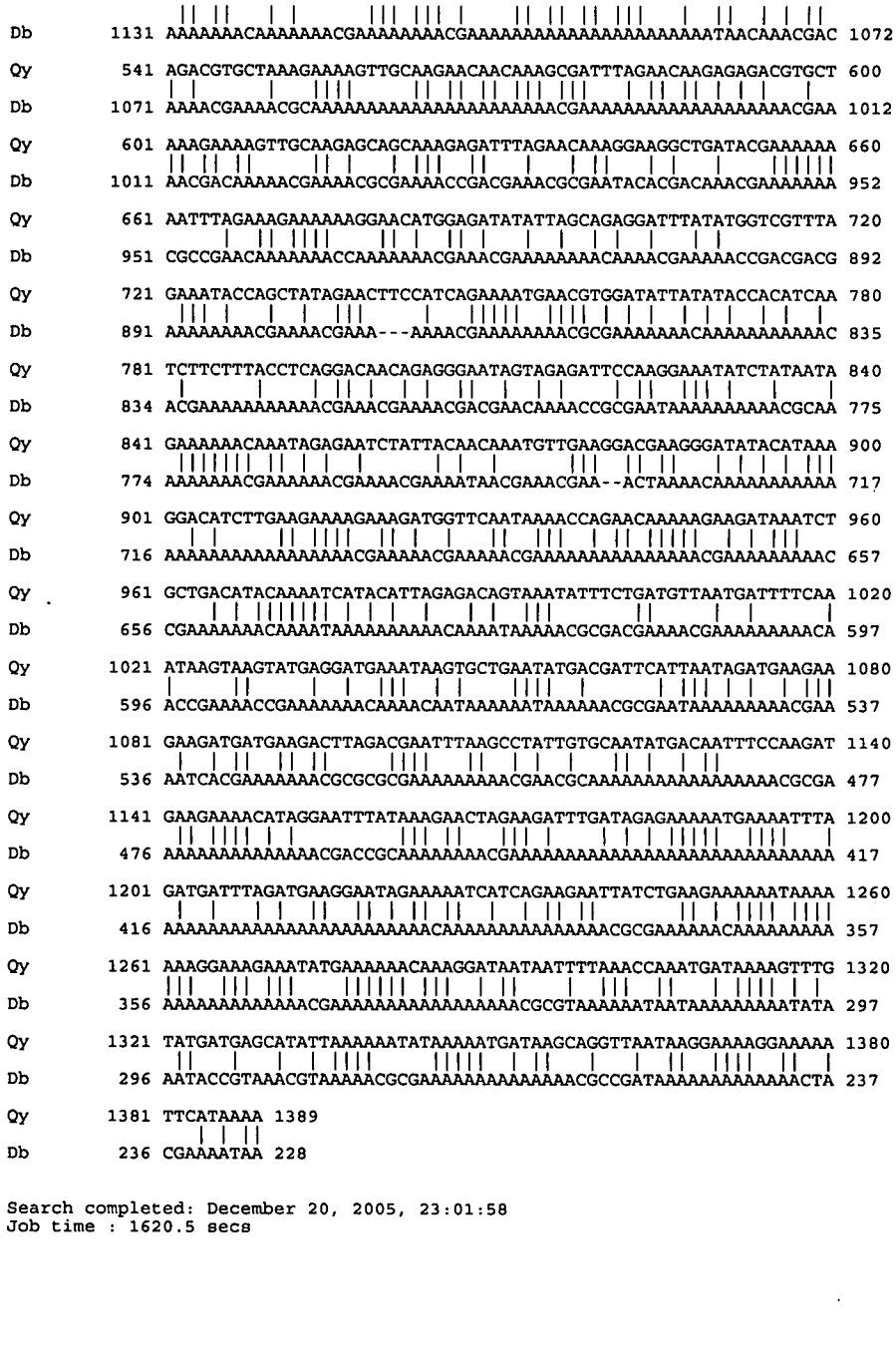
Query Match 8.8%; Score 129.8; DB 8; Length 921;
Best Local Similarity 48.7%; Pred. No. 5.9e-09;
Matches 381; Conservative 0; Mismatches 398; Indels 3; Gaps 1;





RESULT 30
US-10-363-345A-7529/c
; Sequence 7529, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 7529
; LENGTH: 1762
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 7529
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1426, 1496, 1732, 1741..1742, 1745)
US-10-363-345A-7529

Query Match 8.8%; Score 129.8; DB 8; Length 1762;
Best Local Similarity 44.3%; Pred. No. 6.7e-09;
Matches 615; Conservative 0; Mismatches 769; Indels 5; Gaps 2;



Search completed: December 20, 2005, 23:01:58
Job time : 1620.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 15:08:19 ; Search time 408.691 Seconds
(without alignments)
1777.494 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161431 seqs, 245089505 residues

Total number of hits satisfying chosen parameters: 8322862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	77.6	5.2	169495	7	US-11-121-086-61	Sequence 61, Appl
c	2	74.4	5.0	3181	6	US-10-793-626-3866	Sequence 3866, Ap
c	3	73	4.9	3543	6	US-10-793-626-4149	Sequence 4149, Ap
c	4	71.6	4.8	171486	7	US-11-121-086-105	Sequence 105, App
	5	69.6	4.7	2655	6	US-10-793-626-1659	Sequence 1659, Ap
	6	69.6	4.7	3520	6	US-10-793-626-4160	Sequence 4160, Ap
c	7	68.6	4.6	173602	7	US-11-121-086-25	Sequence 25, Appl
	8	68.4	4.6	171486	7	US-11-121-086-105	Sequence 105, App
c	9	67.6	4.6	3602	6	US-10-793-626-3582	Sequence 3582, Ap
	10	67.6	4.6	15071	6	US-10-793-626-2963	Sequence 2963, Ap
	11	67.4	4.5	26772	6	US-10-995-561-13313	Sequence 13313, A
	12	67.4	4.5	54946	6	US-10-995-561-13479	Sequence 13479, A
	13	67.2	4.5	3840	6	US-10-793-626-3187	Sequence 3187, Ap
c	14	67.2	4.5	4965	6	US-10-485-517-43	Sequence 43, Appl
	15	65.6	4.4	119036	6	US-10-995-561-13314	Sequence 13314, A
	16	63.6	4.3	201990	6	US-10-995-561-13303	Sequence 13303, A
c	17	63.2	4.3	1434	9	US-11-082-746-2	Sequence 2, Appli
	18	63.2	4.3	6189	6	US-10-909-125-825	Sequence 825, App
	19	63	4.3	305312	6	US-10-995-561-13236	Sequence 13236, A
	20	62.8	4.2	627	6	US-10-793-626-693	Sequence 693, App
	21	62.8	4.2	627	6	US-10-793-626-1325	Sequence 1325, Ap
c	22	62.8	4.2	2999	6	US-10-793-626-4039	Sequence 4039, Ap
c	23	62.8	4.2	3691	6	US-10-793-626-3338	Sequence 3338, Ap
	24	62.8	4.2	173602	7	US-11-121-086-25	Sequence 25, Appl
	25	61.8	4.2	2925	7	US-11-194-246-84	Sequence 84, Appl
c	26	61.8	4.2	49979	6	US-10-995-561-13443	Sequence 13443, A
	27	61.6	4.2	148935	6	US-10-995-561-13308	Sequence 13308, A
c	28	60.6	4.1	191091	7	US-11-121-086-60	Sequence 60, Appl
	29	59.8	4.0	20317	6	US-10-995-561-13460	Sequence 13460, A
c	30	59.6	4.0	415117	6	US-10-995-561-13274	Sequence 13274, A
c	31	59.4	4.0	3305	6	US-10-793-626-4002	Sequence 4002, Ap
	32	59.4	4.0	150468	7	US-11-112-908-56	Sequence 56, Appl
	33	59.4	4.0	193789	7	US-11-112-908-55	Sequence 55, Appl
c	34	59.2	4.0	2796	6	US-10-793-626-4335	Sequence 4335, Ap
	35	59.2	4.0	394468	6	US-10-995-561-13473	Sequence 13473, A
c	36	58.8	4.0	1267	6	US-10-485-517-36	Sequence 36, Appl
c	37	58.8	4.0	118996	7	US-11-121-086-84	Sequence 84, Appl
	38	58	3.9	1991	7	US-11-196-475-69	Sequence 69, Appl
	39	58	3.9	1991	7	US-11-196-475-77	Sequence 77, Appl
	40	58	3.9	139054	7	US-11-121-086-96	Sequence 96, Appl
	41	57.6	3.9	98862	7	US-11-121-086-76	Sequence 76, Appl
	42	57.6	3.9	182190	7	US-11-121-086-102	Sequence 102, App
	43	57.4	3.9	207908	7	US-11-112-908-21	Sequence 21, Appl
	44	57.4	3.9	212805	7	US-11-112-908-19	Sequence 19, Appl
c	45	57	3.8	151169	7	US-11-121-086-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-11-121-086-61
; Sequence 61, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 169495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (70072)..(70171)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139457)..(157244)
; OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-61

Query Match 5.2%; Score 77.6; DB 7; Length 169495;
Best Local Similarity 45.1%; Pred. No. 0.0057;
Matches 287; Conservative 0; Mismatches 349; Indels 0; Gaps 0;

```
Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              | | || |||      | | | ||  |   ||   |  ||  | ||   || |||||||
Db      67358 ACAGCAGCAACAACCACAAAAGCAGCAATAACTACCACAACGACAGCATCATCAACAACA 67417

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
                 ||  | ||||  ||    |       |  |  |   |||  |||| |||    |
Db      67418 GCAGCATCAACAACAGCATCATCAACAGCATCATCAACAACAGCAGCATCAACAACAGCG 67477

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||| |            | || ||| ||||  |||| |     | | |  |||| |
Db      67478 TCAACAACAAGAGTAACAACCACAACAGTAGCAACCACAAAAGCAGCAACAACCACAACA 67537

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
```

```
Db      67538 GTATCATCAACAGCATCAGCAGCAACCACAACAACAGCAACAGCAACAACCACCACCACT 67597

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302

Db      67598 ACCACAACAAGAGCAACAACTACTACAACAGCAACAACCACCACAACAGAAACGGCAACA 67657

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362

Db      67658 ACCACCACCACCACAACAGCAACAACCACCACAACAGCAACAACCACCACAACAGCAACA 67717

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422

Db      67718 ACTACCACCAACAGCAACAGCATCATCAGCAACAGCAACAACCACCACAAAACAACAGCA 67777

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482

Db      67778 ACAACCACAACAGCAACAGCATCATCAACAGCAACAGCAACAACAACAACGGCAGCAACA 67837

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542

Db      67838 ATAATAGCAACAGCATATTCAACAACAGCAACAACCACATCAACAGCAACAACCACAACC 67897

Qy        543 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 602

Db      67898 ACCTCAAAAACAACAGCAACAACGACAGCAGCAATAACAGCAACAGAAGTGACAACATTA 67957

Qy        603 AGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACA 638

Db      67958 GCAGCAATAACAAAAGCAGCAGTGACAGTAACACCA 67993
```

RESULT 2
US-10-793-626-3866/c
; Sequence 3866, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3866
; LENGTH: 3181
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3866

Query Match 5.0%; Score 74.4; DB 6; Length 3181;
Best Local Similarity 48.7%; Pred. No. 0.015;
Matches 266; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

```
Qy        873 TGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAAT 932

Db        877 TATTAAAGAACATCCAGATGTACAGAAAGTAACTAATGCTACATCTAAAGCTACTTCAAA 818

Qy        933 AAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGT 992

Db        817 AACAACTTCTGCAATTTCAAATAAAGCTTCAGATGTCAAAGAGTACGTAGGAGATAAAAA 758

Qy        993 AAATATTTCTGATGTTAA---TGATTTTCAAATAAGTAAGTATGAGGATGAAATA----- 1044

Db        757 GCAAGAATTTGATAATAAGCGTGAATTGAAGAAAATTGCGAAGGAAAATGATCCAGCCTA 698

Qy       1045 -AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGA 1103
```

```
Db        697 CATTGAGAAAAAAGGCGAAAAATTAGCTAAACAAAATCGTAAAGAAGCAGACAAAATGGA 638

Qy       1104 ATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAA 1163

Db        637 CAAAAAGCTTCAAAAGAATATTGAGAAACGTCATAAAGAAGAACAAAAACAACGTGAAGC 578

Qy       1164 AGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGA 1223

Db        577 TAATCAGAAACAACGTATTAAAGACATGAAAAAAACTCAAAAATACGAAGAGCAAGTTGG 518

Qy       1224 AAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAAC 1283

Db        517 CTTAACTCCTGGTAAAATAGATCACGAAATTGAGAAAAAAGGCGAAAAACTAGAAAAAGA 458

Qy       1284 AAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATA 1343

Db        457 TAATCGTAAAGATATTAAAAAATTAGATAAAAAGCTTCAAAAGAATATTGAAAAACGTCA 398

Qy       1344 TAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATAT 1403

Db        397 TAAAGAAGAACAAAAACAGCGTGAAGAAGCAGAGAAAGCTAGAAAAAAAGAATTTAAAAA 338

Qy       1404 ATTTGA 1409

Db        337 ATATGA 332
```

RESULT 3
US-10-793-626-4149/c
; Sequence 4149, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4149
; LENGTH: 3543
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4149

Query Match 4.9%; Score 73; DB 6; Length 3543;
Best Local Similarity 43.1%; Pred. No. 0.023;
Matches 581; Conservative 0; Mismatches 750; Indels 18; Gaps 4;

```
Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70

Db       3441 AAAGAAATGCTAAAAATGATATTGATATTAAAGCAGCTGATAAGAAAATAAAAATTCAAA 3382

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130

Db       3381 GAATAAATGATGCTACAGATGAAGAAATTCAAGAAGCGAATCGTAAAATTGAAGAAGCTA 3322

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190

Db       3321 AGATTGAAGCAAAAGATAATATTCA-----ACGCAATAGTACTAGAGATCAAGTAAATGA 3267

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250

Db       3266 AGCGAAAACTAATGGAATAAATAAAATAGAAAATATAACACCAGCAACTACTGTGAAATC 3207

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310

Db       3206 TGAAGCTAGACAAGCAGTACAGAATAAAGCAA-ATGAACAGATTAATCATATTCAAAACA 3148
```

```
Qy       311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
Db      3147 CGCCTGATGCAACTAATGAAGAAAAACAAGAGGCAATAAATAGAGTAAGTGCTGAATTAG 3088

Qy       371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
Db      3087 CAAGAGTTCAAGCACAAATAAATGCAGAACATACAACCCAAGGTGTCAAAACTATCAAAG 3028

Qy       431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
Db      3027 ACGACGCGATAACTTCTTTATCTCGAATTAATGCACAAGTTGTTGAGAAAGAGTCTGCAA 2968

Qy       491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
Db      2967 GAAATGCAATCGAACAAAAGGCAACACAACAAACGCAATTTATTAATAATA---ATGATA 2911

Qy       551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
Db      2910 ATGCTACAGATGAAGAAAAAGAGGTCGCCAACAATTTAGTTATCGCTACAAAACAAAAAT 2851

Qy       611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
Db      2850 CATTAGATAATATTAACTCCTTATCTTCAAATAATGATGTTGAAAATGCTAAAGTAGCAG 2791

Qy       671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
Db      2790 GAATAAATGAAATAGCTAACGTTTTACCAGCAACCGCTGTTAAGTCAAAAGCAAAAAAAG 2731

Qy       731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
Db      2730 ATATTGATCAAAAACTCGCGCAACAGATTAATCAAATTCAAACGCATCAAACTGCTACAA 2671

Qy       791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 850
Db      2670 CTGAGGAAAAAGAAGCGGCTATTCAATTGGCAAATCAAAAATCAAATGAAGCAAGAACAG 2611

Qy       851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
Db      2610 CAATTCAAAATGAACATAGTAACAATGGTGTCGCACAAGCTAAATCTAACGGCATTCATG 2551

Qy       911 A---------AGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db      2550 AAATTGAATTAGTTATGCCAGATGCGCACAAAAAATCTGATGCTAAACAAAGTATCGATA 2491

Qy       962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db      2490 ATAAATATAATGAGCAAAGTAATACTATCAACACTACACCAGATGCAACAGATGAAGAAA 2431

Qy      1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db      2430 AGCAAAAAGCATTAGATAAATTAAAAATAGCTAAAGATGCAGGATACAACAAAGTTGATC 2371

Qy      1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
Db      2370 AAGCGCAAACAAACCAACAAGTATCTGATGCAAAAACTGAGGCTATAGATACGATAACTA 2311

Qy      1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
Db      2310 ATATTCAAGCAAATGTTGCAAAAAAACCATCCGCTCGAGTGGAATTAGATTCAAAGTTTG 2251

Qy      1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db      2250 AGGATTTAAAGCGTCAAATCAATGCAACGCCCAATGCTACAGAAGAAGAAAAACAAGATG 2191

Qy      1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
Db      2190 CAATTCAAAGATTGAATGGTAAAAGAGATGAAGTTAAGAATCTAATAAATCAAGATAGAC 2131

Qy      1322 ATGATGAGCATATTAAAAAATATAAAAAT 1350
Db      2130 GTGACAATGAAGTTGAACAGCACAAAAAT 2102
```

```
RESULT 4
US-11-121-086-105/c
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
;   LENGTH: 171486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-105

  Query Match             4.8%;  Score 71.6;  DB 7;  Length 171486;
  Best Local Similarity   45.7%;  Pred. No. 0.037;
  Matches  439;  Conservative    0;  Mismatches  509;  Indels   12;  Gaps    5;

Qy       530 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 589
Db     17309 TATAAATATATAAATATATTTATATATATCTATAATATATAAATATATAAATATATTTAT 17250

Qy       590 AGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTG 649
Db     17249 ATATCTATAATATATAAATATATAAATATATTTATATATATCTATAATATATAAATATAT 17190

Qy       650 ATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTAT 709
Db     17189 AAATATATTTATATATATCTATAATATATAAATATATAAATATATTTATATATCTATAAT 17130

Qy       710 ATGGTCGTTTAGAAATA-----CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATA 764
Db     17129 ATATAAATATATAAATATATTTATATATATCTATAATATATAAATATATAAATTTATATA 17070

Qy       765 TTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAA 824
Db     17069 TATCTATAATATATAAATATATAAATATATCTATCATATATAAATATACAAATATATA 17010

Qy       825 GGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACG 884
Db     17009 TCTATCATATATAAATATACAAATATATTTATATATATCTATCATATATAAATATACAAA 16950

Qy       885 AAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACA 944
Db     16949 TATATTTATATATATCTATCATATATAAATATACAAATATATTTATATATATCTATCATA 16890

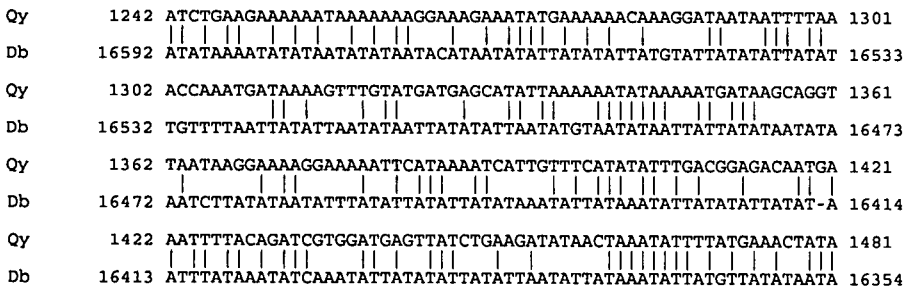
Qy       945 AAAAGAAGATAAATCTGCTGACATACAAA---ATCATACATTAGAGACAGTAAATATTTC 1001
Db     16889 TATAATATATAAATGTATTTATATATAAATATATAATATATTATATATTATAGATAATAT 16830

Qy      1002 TGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGA 1061
Db     16829 ATGTATTATCTATAATTATATAATATAATTATATATAATATGTAATGTATTATATGTTAT 16770

Qy      1062 TTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCA 1121
Db     16769 ATATTTAAGACATATATAATATATT--ATATTAATTAATATATTTATTATATAATATATT 16712

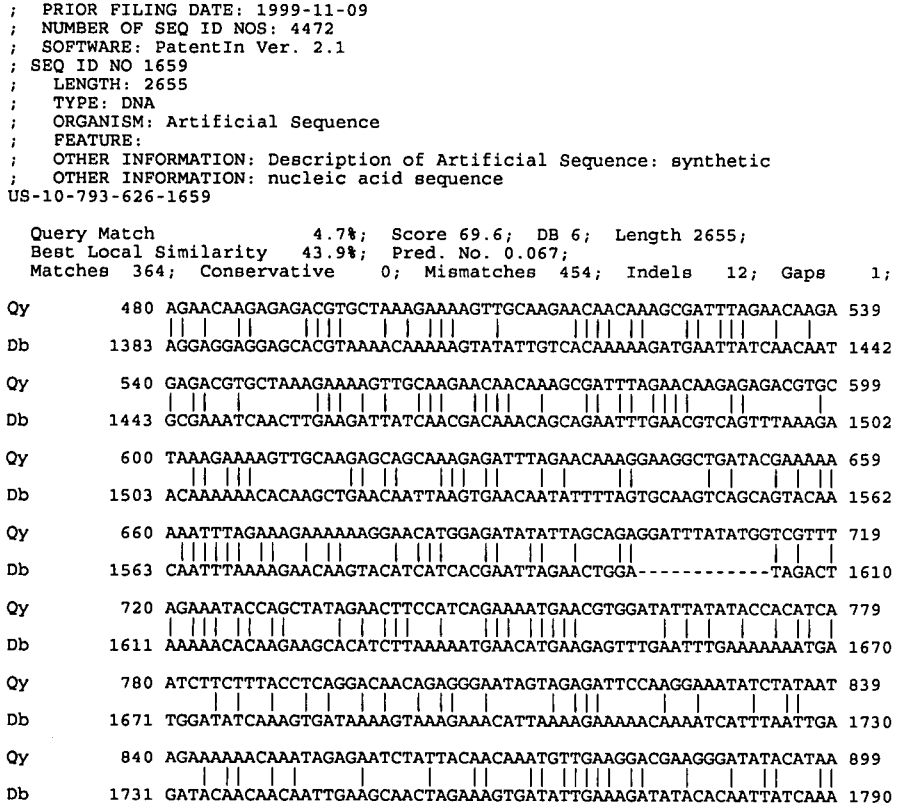
Qy      1122 ATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGAT 1181
Db     16711 ATATTAATATATTTATTATATAATTAATATATTTA-TTATATTAATATGTAATATATTAT 16653

Qy      1182 AGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATT 1241
Db     16652 ACTATATAATAAATATATATAATATATTTATTAATATATTATATAATAATATATAATATT 16593
```



RESULT 5
US-10-793-626-1659
; Sequence 1659, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1659
; LENGTH: 2655
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1659

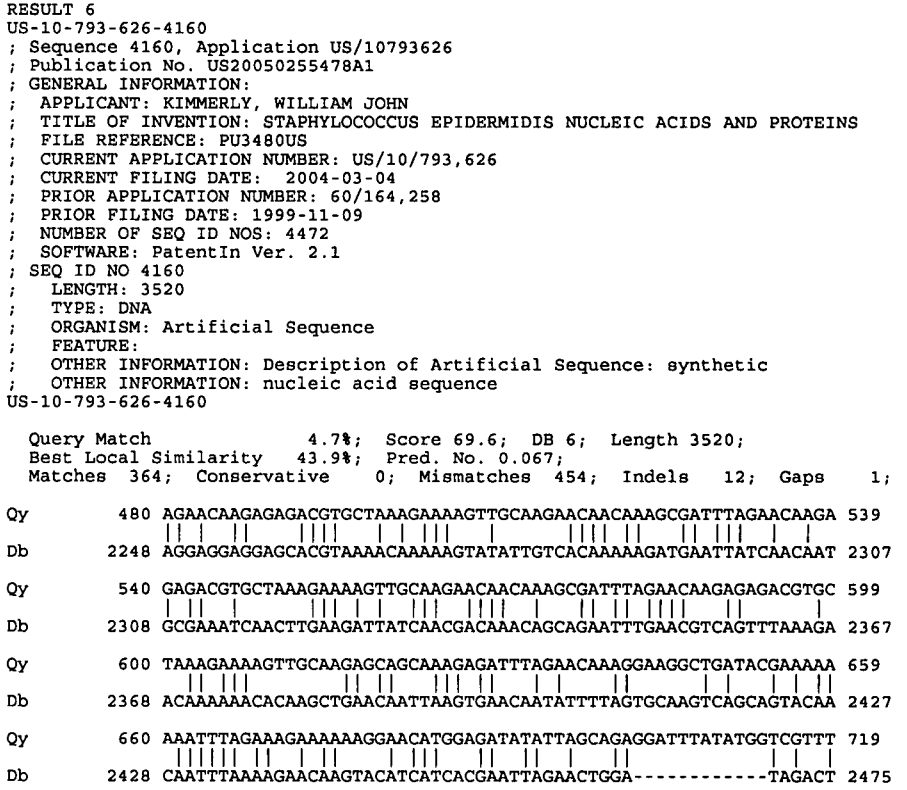
Query Match 4.7%; Score 69.6; DB 6; Length 2655;
Best Local Similarity 43.9%; Pred. No. 0.067;
Matches 364; Conservative 0; Mismatches 454; Indels 12; Gaps 1;





RESULT 6
US-10-793-626-4160
; Sequence 4160, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4160
; LENGTH: 3520
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4160

Query Match 4.7%; Score 69.6; DB 6; Length 3520;
Best Local Similarity 43.9%; Pred. No. 0.067;
Matches 364; Conservative 0; Mismatches 454; Indels 12; Gaps 1;



```
Db       2476 AAAAACACAAGAAGCACATCTTAAAAATGAACATGAAGAGTTTGAATTTGAAAAAAATGA 2535

Qy        780 ATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAAT 839

Db       2536 TGGATATCAAAGTGATAAAAGTAAAGAAACATTAAAAGAAAAACAAAATCATTTAATTGA 2595

Qy        840 AGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAA 899

Db       2596 GATACAACAACAATTGAAGCAACTAGAAAGTGATATTGAAAGATATACACAATTATCAAA 2655

Qy        900 AGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATC 959

Db       2656 AGAAGGAAAAGCTTCGACACATCAAACACAACAACAACTACATCAAAAACAATCTGATTT 2715

Qy        960 TGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCA 1019

Db       2716 AGCTGTTGTTAAAGAGCGAATTAAATCGCAAAAGCAAGTTTATGAACGTTTAGATAAACA 2775

Qy       1020 AATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGA 1079

Db       2776 ACTTAGCGATTCAGAACGTCAAAAAATTGAAGTAAATGAAAAAATCAAATTGTTTAATTC 2835

Qy       1080 AGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGA 1139

Db       2836 AGATGAAATGATGGGTAAAGATGCTTTTGAAAAGTTGAAAGAGCAAATTCAGCAACAAGA 2895

Qy       1140 TGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTT 1199

Db       2896 AAATGTAAGACAAAATTTAAATCAACAACTTAGTGAGATTAAACAGCAACGTAAAGATCT 2955

Qy       1200 AGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAA 1259

Db       2956 TAATGAGAAAATCGAAATAAATGAAAGTCAGCTTCAAAAATGTCATCAAGATATACTTTC 3015

Qy       1260 AAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG 1309

Db       3016 TATAGAAAATCATTATCAAGATATTAAAGCAAAACAATCAAAGCTAGATG 3065
```

RESULT 7
US-11-121-086-25/c
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25

Query Match 4.6%; Score 68.6; DB 7; Length 173602;
Best Local Similarity 44.5%; Pred. No. 0.092;
Matches 365; Conservative 0; Mismatches 449; Indels 7; Gaps 2;

```
Qy        656 AAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTC 715

Db     136831 AATAATATATAAACATATTATATAAATATAATAATATATAAACATATTATATAAATATAA 136772

Qy        716 GTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCAC 775

Db     136771 TAATATATAAACATATTATATAAATATAATATATAAATATTATATAAATATATATATAAT 136712

Qy        776 ATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTA 835

Db     136711 ATATAAAATATAT-----TATATATAATATACTATATAATATAGCATATAAATATATATA 136657

Qy        836 TAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATAC 895

Db     136656 ATATATTATATAAAAATATATAACATATTATATAAAATATTATATAAATATATAATAT-- 136599

Qy        896 ATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATA 955

Db     136598 ATAAATATATTTATATATTATATATTTATATATACATAAACATATAAATATATAATATAT 136539

Qy        956 AATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATT 1015

Db     136538 ACTATATAAATATATTATATAAATATATATAAATATATACTATATTTATAATATAGTATA 136479

Qy       1016 TTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATG 1075

Db     136478 TATTTATATAGATTTATATAATTTATATATATTTTATATATTATATATAATTTATATATA 136419

Qy       1076 AAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCC 1135

Db     136418 TTTTATATATTATATATAATTTATATATATTTATATATTATATATAATTTATATATATTG 136359

Qy       1136 AAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAA 1195

Db     136358 TATATATTATATATAAAATATATAATTTATATATAATATATAAATTATATAATTTATATA 136299

Qy       1196 ATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAA 1255

Db     136298 TATTTATATATTATATATAAATATATATAATTTATATAATTTATATATTATATATATAAA 136239

Qy       1256 TAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAA 1315

Db     136238 TATATAATTTATATATATTTATATAATATATATTATATATTTATATATATAATATATATT 136179

Qy       1316 GTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGG 1375

Db     136178 TATATTTATATATCTATATAATATATATTATATATTTATATATTTATAAATATATAATAT 136119

Qy       1376 AAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCG 1435

Db     136118 ATATTTATATATAAAAATATATACATATATAATTTATATATATATTACATATACATATAA 136059

Qy       1436 TGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAA 1476

Db     136058 AATATATGTAAAATGTATGTGTATGAATATGTATATATAAA 136018
```

RESULT 8
US-11-121-086-105
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 171486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-105

Query Match 4.6%; Score 68.4; DB 7; Length 171486;
Best Local Similarity 46.2%; Pred. No. 0.098;
Matches 370; Conservative 0; Mismatches 421; Indels 9; Gaps 4;

```
Qy        681 ACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACT 740
              | ||  | |||||  ||  | |   | |||||  |   | | | |||  |  ||||  | |
Db      16086 AAATATAAATATAATATTATATATTATATATATTGTATAATATATATAATATATTA-T 16144

Qy        741 TCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAA 800
                ||   | |   | | |  ||||||||||  | || |   |  |   |  | |   |
Db      16145 ATTATATAATATATATCTTATATATTATATATTATATAAGATATAATATTATATGTTATA 16204

Qy        801 CAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATC 860
              |      ||  || |   |    |  |||  ||| |||    |     | |||   |||
Db      16205 TATTATATATTATATAATATATTATATTATAATTATTATAATTATGTTATATAATTAATA 16264

Qy        861 TATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAA 920
              |  || || | ||  |             ||||  ||| |  | ||  |  | || | |
Db      16265 TTATATAATATATAATATGT------ATAATATTAATATAATATATAATATATAACATAT 16318

Qy        921 AGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATAC 980
              |     | | ||| |  | |   |  | | |||| ||    | |||||  |  |  |
Db      16319 AATATTTATATTAATATAATATATATTATATGATATATTATATAACATAATATTTATAAT 16378

Qy        981 ATTAGAGACAGTAAATATTTCTGATGTTAAT-GATTTTCAAATAAGTAAGTATGAGGATG 1039
              ||||      |   | || | |||| || ||  ||| |  ||||  |||   | |  ||
Db      16379 ATTAATATAATATATAATATTTGATATTTATAAATTATATAATATATAATATTTATAATA 16438

Qy       1040 AAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAG 1099
                |     ||||| |  |||   ||| |  |  |  | |  || || |      |
Db      16439 TTTATATAATATAATATAAATATTATATAAGATTTATATTATATAATAATTATATTACAT 16498

Qy       1100 ACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTT 1159
              |  ||| ||    |||  |   |||    ||     |  ||  | || |||||  | |
Db      16499 ATTAATATATAATTATATTAATATAATTAAAACAATATAATATATAATACATAATATATA 16558

Qy       1160 ATA-AAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGA 1218
              ||| |  |  ||  | || | ||| |    |   || |||  |     ||| || |   |
Db      16559 ATATATTATGTATTATATATTATATATTTTATATAATATTATATATTATTATATAATATA 16618

Qy       1219 ATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAA 1278
              ||  |||   || | |   |||   | | | | | | |   |  ||       |
Db      16619 TTAATAAATATATTATATATATTTATTATATAGTATAATATATTACATATTAATATAATA 16678

Qy       1279 AAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
              || | |     ||   | |  | |    | |||| |  || ||| || |  |||||||
Db      16679 AATATATTAATTATATAATAAATATATTAATATAATATATTATATAATAAATATATTTAAT 16738

Qy       1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
              |||||||  |      |  | | |   | ||   | ||| ||| | ||  |   |
Db      16739 TAATATAATATATTATATATGTCTTAAATATATAACATATAATACATTACATATTATATA 16798

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
              |  ||| | |   |   |  || | |  | | ||  |  || |  |||   |   ||||
Db      16799 TAATTATATTATATAATTATAGATAATACATATATTATCTATAATATATAATATATTATA 16858

Qy       1459 ACTAAATATTTTATGAAACT 1478
                |  |||| | |  | | |
Db      16859 TATTTATATATAAATACATT 16878


RESULT 9
US-10-793-626-3582/c
; Sequence 3582, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3582
;  LENGTH: 3602
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: synthetic
;  OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3582

  Query Match             4.6%;  Score 67.6;  DB 6;  Length 3602;
  Best Local Similarity   43.1%;  Pred. No. 0.12;
  Matches  438;  Conservative    0;  Mismatches  569;  Indels    9;  Gaps    2;

Qy        333 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACG 392
              ||  |  | || ||||| ||  ||  |  |   | |    ||||   |  | |  | |
Db      2208 AGTAAATCCTGTTAAAGCAAGTTTAAATTATAGAAACGCTGATTATAATTTAAAACGTCA 2149

Qy        393 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGA 452
              |||  ||   ||  |||||    || |||  ||| |  || | ||   | | || |
Db      2148 ATTTAACAAAGCTTTAAAAGAAGCTAAAGGCGTATTAAATAAAAATAGCGGTACAAATGT 2089

Qy        453 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 512
              ||  |  | ||    |||     ||||  ||||        |||||||||  | ||  |
Db      2088 CAATATCAATGACATTCAACATCTTTTAACACAAATAGATAATGCTAAAGACCAATTAAA 2029

Qy        513 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 572
              |    ||| | |       ||||||  |     |   || ||| |      |    || ||
Db      2028 TGGTGAACGACGTCTAAAAGAACATCAACAAAAATCTGAAGTATTTATTATTAAAGAATT 1969

Qy        573 AAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTT 632
              |       |  |  | |     |  |||  |  ||          | |||   ||||  |
Db      1968 AGATATACTTAATAATGCTCAAAAAGCTGCAATAA------TTAATCAGATTAGAGCGTC 1915

Qy        633 AGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATAT 692
              || | |  ||     ||   |||    |   |  || | |   | |||  | |||
Db      1914 TAAAGACATTAAAATAATTAATCAAATCGTTGATAATGCAATAGAATTAAATGATGCTAT 1855

Qy        693 ATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAA 752
              ||       |  | |||   | |  | || | || ||| | ||          ||| |
Db      1854 GCAAGGTTTAAAAGAACATGTAGCTCAATTAACAGCAACTACAAAAGACAACATTGAATA 1795

Qy        753 TGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAG 812
              |  |  ||   || |    |   |   |  | |              ||||  || ||||||
Db      1794 TTTAAATGCTGATGAAGACCTTAAAATACAATATGATTACGCTATCAACTTAGCGAATAA 1735

Qy        813 TAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAA 872
              |    |  ||| ||||    || ||  || |  | ||||    |  |   |  |  |
Db      1734 TGTTCTTGACAAAGAAAACGGTACAAATAAAGACGCTAATATCATAATTGGAATGATTCA 1675

Qy        873 TGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAAT 932
                  |                 || |  |     ||||  |  |||||||  |  ||
Db      1674 AAACATGGATGATGCTAGAGCACTTCTAAATGGAATTGAAAGACTTAAAGATGCTCAAAC 1615

Qy        933 AAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGT 992
              |||| || |  |  |      |  |  |      |    |||  | ||  | | ||    |
Db      1614 AAAAGCACATAATGACATTAAAGATACGCTCAAACGTCAACTTGATGAAATTGAACACGC 1555

Qy        993 AAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGA 1052
              |||    |    |  | |  ||||  ||  |   |  ||||  |  |||
Db      1554 TAATGCAACATCAAATTCTAAAGCTCAAGCAAAACAAATGGTAAATGAGGAA---GCTAG 1498

Qy        1053 ATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCC 1112
               | | |    |||     ||  || | | |    ||||      |  |  ||    ||
Db      1497 AAAAGCGTTTTCTAATATTAATCACGCAACATCAAATGATTTAGTTAATCAAGCAAAAGA 1438

Qy        1113 TATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGA 1172
               |   | ||||| ||  | |   ||      | | | |   |  |  | |   | |||||||
Db      1437 TGAAGGGCAATCTGCAATTGAACACATACATGCAGATGAATTACCTAAAGCAAAACTAGA 1378

Qy        1173 AGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATC 1232
               | |     |  ||   | ||||  | || ||| || || |   |||     ||     |
Db      1377 TGCTAATCAAATGATTGACCAAAAAGTTGAAGATATAAATCACTTAATTAGTCAAAATCC 1318

Qy        1233 AGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAA 1292
```

```
Db      1317 AAACTTATCAAATGAAGAAAAAAATAAACTAATATCTCAAATTAATAAGTTAGTAAATGG 1258

Qy      1293 TAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348

Db      1257 AATTAAGAATGAAATTCAACAAGCTATAAACAAACAACAAATAGAAAATGCTACAA 1202


RESULT 10
US-10-793-626-2963
; Sequence 2963, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2963
;   LENGTH: 15071
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2963

  Query Match             4.6%;  Score 67.6;  DB 6;  Length 15071;
  Best Local Similarity   43.1%;  Pred. No. 0.12;
  Matches  438;  Conservative    0;  Mismatches  569;  Indels    9;  Gaps    2;

Qy       333 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACG 392

Db     12864 AGTAAATCCTGTTAAAGCAAGTTTAAATTATAGAAACGCTGATTATAATTTAAAACGTCA 12923

Qy       393 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGA 452

Db     12924 ATTTAACAAAGCTTTAAAAGAAGCTAAAGGCGTATTAAATAAAAATAGCGGTACAAATGT 12983

Qy       453 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 512

Db     12984 CAATATCAATGACATTCAACATCTTTTAACACAAATAGATAATGCTAAAGACCAATTAAA 13043

Qy       513 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 572

Db     13044 TGGTGAACGACGTCTAAAAGAACATCAACAAAAATCTGAAGTATTTATTATTAAAGAATT 13103

Qy       573 AAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTT 632

Db     13104 AGATATACTTAATAATGCTCAAAAAGCTGCAATAA------TTAATCAGATTAGAGCGTC 13157

Qy       633 AGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATAT 692

Db     13158 TAAAGACATTAAAATAATTAATCAAATCGTTGATAATGCAATAGAATTAAATGATGCTAT 13217

Qy       693 ATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAA 752

Db     13218 GCAAGGTTTAAAAGAACATGTAGCTCAATTAACAGCAACTACAAAAGACAACATTGAATA 13277

Qy       753 TGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAG 812

Db     13278 TTTAAATGCTGATGAAGACCTTAAAATACAATATGATTACGCTATCAACTTAGCGAATAA 13337

Qy       813 TAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAA 872

Db     13338 TGTTCTTGACAAAGAAAACGGTACAAATAAAGACGCTAATATCATAATTGGAATGATTCA 13397

Qy       873 TGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAAT 932

Db     13398 AAACATGGATGATGCTAGAGCACTTCTAAATGGAATTGAAAGACTTAAAGATGCTCAAAC 13457

Qy       933 AAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGT 992

Db     13458 AAAAGCACATAATGACATTAAAGATACGCTCAAACGTCAACTTGATGAAATTGAACACGC 13517

Qy       993 AAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGA 1052

Db     13518 TAATGCAACATCAAATTCTAAAGCTCAAGCAAAACAAATGGTAAATGAGGAA---GCTAG 13574

Qy      1053 ATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCC 1112

Db     13575 AAAAGCGTTTTCTAATATTAATCACGCAACATCAAATGATTTAGTTAATCAAGCAAAAGA 13634

Qy      1113 TATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGA 1172

Db     13635 TGAAGGGCAATCTGCAATTGAACACATACATGCAGATGAATTACCTAAAGCAAAACTAGA 13694

Qy      1173 AGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATC 1232

Db     13695 TGCTAATCAAATGATTGACCAAAAAGTTGAAGATATAAATCACTTAATTAGTCAAAATCC 13754

Qy      1233 AGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAA 1292

Db     13755 AAACTTATCAAATGAAGAAAAAAATAAACTAATATCTCAAATTAATAAGTTAGTAAATGG 13814

Qy      1293 TAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348

Db     13815 AATTAAGAATGAAATTCAACAAGCTATAAACAAACAACAAATAGAAAATGCTACAA 13870


RESULT 11
US-10-995-561-13313
; Sequence 13313, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13313
;   LENGTH: 26772
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13313

  Query Match             4.5%;  Score 67.4;  DB 6;  Length 26772;
  Best Local Similarity   46.1%;  Pred. No. 0.13;
  Matches  323;  Conservative    2;  Mismatches  373;  Indels    3;  Gaps    3;

Qy       691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db     21231 ATTATATATAATTATATAAAATATAATATAGAATATCTAATAATGTATAATATATAACAT 21290

Qy       751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810

Db     21291 ATAAAATAATATTATTTAATATATAATATTTTATATATAATATTTTTATATATAATATAA 21350

Qy       811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870

Db     21351 TATATATTTTATATAATTATTAATTATATAATAATATATAATATATTTTATACAT 21410

Qy       871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930

Db     21411 AATTATTAATTA-TATATAATTAATATATAATATATCTTATACATAATTATCAATTATAT 21469

Qy       931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
```

```
Db      21470 ATAATTAATATATAATATATATTTTATACATAATTATTAATTATATATAATTAATATATA 21529

Qy        991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA-AGTGC 1049

Db      21530 ATATATCTTATACATAATATATATAAATATATTATATATAATATATATTATATATAATAT 21589

Qy       1050 TGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAA 1109

Db      21590 TATATATAATATATATTATATATATATAAAATTTATATATAATATTATATATAATATTATAT 21649

Qy       1110 GCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACT 1169

Db      21650 ATTTTATATACAATATGATATATAATATAATTTATATATTATATATATTTATATATAATT 21709

Qy       1170 AGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATC 1229

Db      21710 ATTATATAAATTATATAAATATAAATTATATATTTATATATAATTATTATATAAATCATT 21769

Qy       1230 ATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGA 1289

Db      21770 AT-ATAATTATTATAATTATAATATATAATATAATATAATATTATATATAMTATATAGTA 21828

Qy       1290 TAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAA 1349

Db      21829 TTCTATATAAATAATATAACATATATTTTATATAGAATATTATATATAATATAATATATA 21888

Qy       1350 TGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAAT 1390

Db      21889 TTTTATATAGAATATTATATATAATATAATATAYATTTTAT 21929


RESULT 12
US-10-995-561-13479
; Sequence 13479, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13479
;   LENGTH: 54946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13479

  Query Match             4.5%;  Score 67.4;  DB 6;  Length 54946;
  Best Local Similarity   46.1%;  Pred. No. 0.13;
  Matches  323;  Conservative    2;  Mismatches  373;  Indels    3;  Gaps    3;

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db       2685 ATTATATATAATTATATAAAATATAATATAGAATATCTAATAATGTATAATATATAACAT 2744

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810

Db       2745 ATAAAATAATATTATTTAATATATAATATTTTATATATAATATTTTTATATATAATATAA 2804

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870

Db       2805 TATATATTTTATATATAATTATTAATTATATAATTAATATATAATATATTTTATACAT 2864

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930

Db       2865 AATTATTAATTA-TATATAATTAATATATAATATATCTTATACATAATTATCAATTATAT 2923

Qy        931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990

Db       2924 ATAATTAATATATAATATATATTTTATACATAATTATTAATTATATATAATTAATATATA 2983

Qy        991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA-AGTGC 1049

Db       2984 ATATATCTTATACATAATATATATAAATATATTATATATAATATATATTATATATAATAT 3043

Qy       1050 TGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAA 1109

Db       3044 TATATATAATATATATTATATATATAAAATTTATATATAATATTATATATAATATTATAT 3103

Qy       1110 GCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACT 1169

Db       3104 ATTTTATATACAATATGATATATAATATAATTTATATATTATATATATTTATATATAATT 3163

Qy       1170 AGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATC 1229

Db       3164 ATTATATAAATTATATAAATATAAATTATATATTTATATATAATTATTATATAAATCATT 3223

Qy       1230 ATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGA 1289

Db       3224 AT-ATAATTATTATAATTATAATATATAATATAATATAATATTATATATAMTATATAGTA 3282

Qy       1290 TAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAA 1349

Db       3283 TTCTATATAAATAATATAACATATATTTTATATAGAATATTATATATAATATAATATATA 3342

Qy       1350 TGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAAT 1390

Db       3343 TTTTATATAGAATATTATATATAATATAATATAYATTTTAT 3383


RESULT 13
US-10-793-626-3187
; Sequence 3187, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3187
;   LENGTH: 3840
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3187

  Query Match             4.5%;  Score 67.2;  DB 6;  Length 3840;
  Best Local Similarity   43.4%;  Pred. No. 0.14;
  Matches  425;  Conservative    0;  Mismatches  543;  Indels   12;  Gaps    2;

Qy        380 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 439

Db        137 AAGCACAAATAAATGCAGAACATACAACCCAAGGTGTCAAAACTATCAAAGACGACGCGA 196

Qy        440 GACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 499

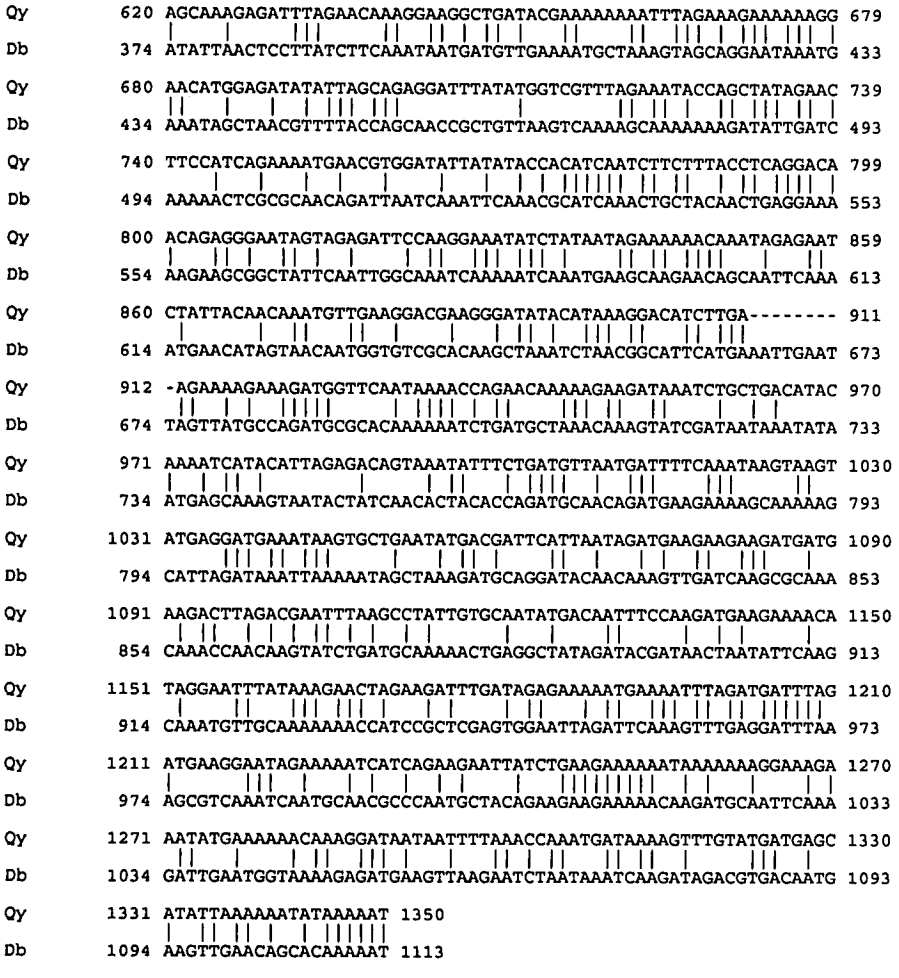
Db        197 TAACTTCTTTATCTCGAATTAATGCACAAGTTGTTGAGAAAGAGTCTGCAAGAAATGCAA 256

Qy        500 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 559

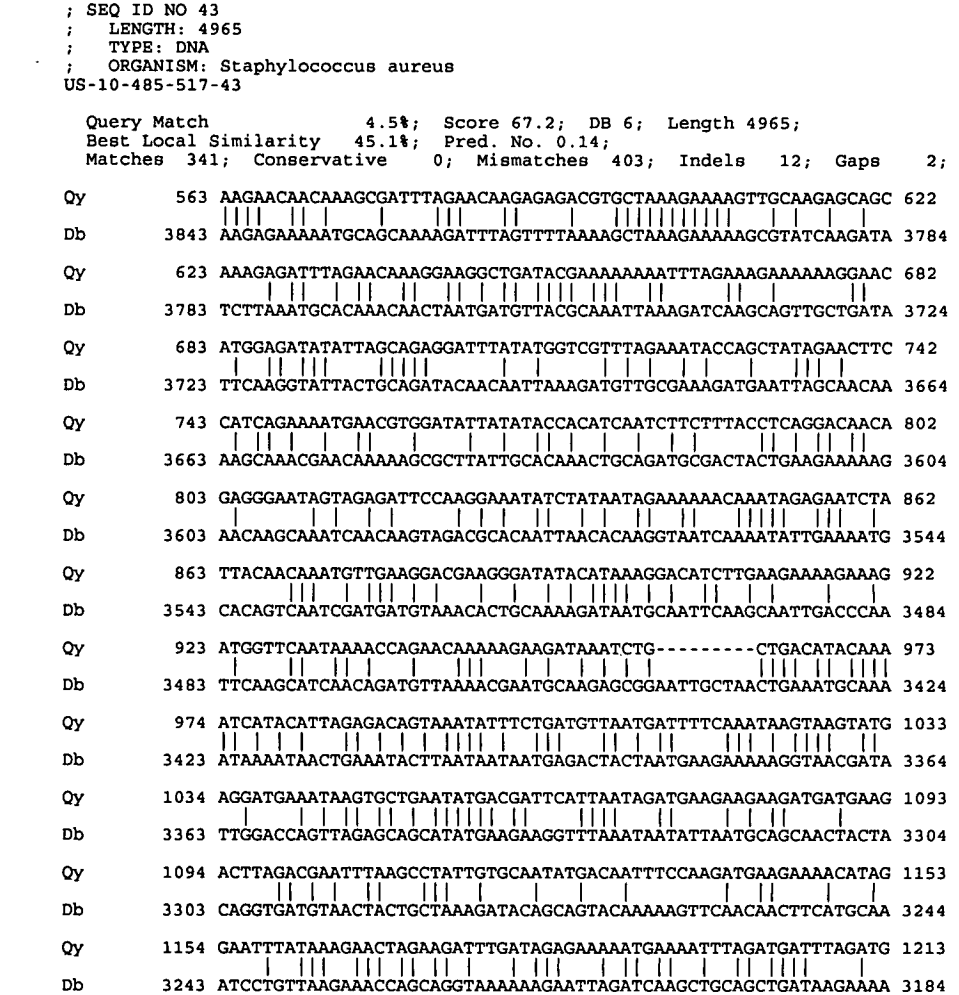
Db        257 TCGAACAAAAGGCAACACAACAAACGCAATTTATTAATAATA---ATGATAATGCTACAG 313

Qy        560 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGC 619

Db        314 ATGAAGAAAAAGAGGTCGCCAACAATTTAGTTATCGCTACAAAACAAAAATCATTAGATA 373
```



```
RESULT 14
US-10-485-517-43/c
; Sequence 43, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
;  APPLICANT: University of Sheffield
;  APPLICANT:  Biosynexus Incorporated
;  APPLICANT:  Foster, Simon
;  APPLICANT:  Mond, James
;  TITLE OF INVENTION: Antigenic Polypeptides
;  FILE REFERENCE: P100629WO
;  CURRENT APPLICATION NUMBER: US/10/485,517
;  CURRENT FILING DATE:  2004-02-02
;  PRIOR APPLICATION NUMBER: GB 0118825.9
;  PRIOR FILING DATE: 2001-08-02
;  PRIOR APPLICATION NUMBER: GB 0200349.9
;  PRIOR FILING DATE: 2002-01-09
;  NUMBER OF SEQ ID NOS: 424
;  SOFTWARE: PatentIn version 3.1
```



```
; SEQ ID NO 43
;   LENGTH: 4965
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-485-517-43

  Query Match             4.5%;  Score 67.2;  DB 6;  Length 4965;
  Best Local Similarity   45.1%;  Pred. No. 0.14;
  Matches  341;  Conservative    0;  Mismatches  403;  Indels   12;  Gaps    2;
```



```
RESULT 15
US-10-995-561-13314
; Sequence 13314, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
```

```
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13314
;   LENGTH: 119036
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(119036)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13314

  Query Match             4.4%;  Score 65.6;  DB 6;  Length 119036;
  Best Local Similarity   45.6%;  Pred. No. 0.23;
  Matches  304;  Conservative    1;  Mismatches  355;  Indels    6;  Gaps    2;

Qy        814 AGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
              | || ||  ||| ||| ||      |   |||||    |  |   | ||||   | ||||
Db      31214 AAAGTTTAAAAGAAAAAATTATATTTTAGAAAAATGGCTGTATTTTTTATTTAGAAAAAT 31273

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
              |       |   | |  | || ||| |    |  |  |   |||  |||||  ||  |||
Db      31274 GGCTGTATAAATAAGATTTTATATATATATAAAATATATATAAAATAAGAT-TTTATATA 31332

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTA 993
              |  | || | | || ||||   |    | ||| | ||       || |   ||
Db      31333 TATATAAAATATATATAAAATAAGATTTTATATATATATAAAATATATATAAAATAAGAT 31392

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
              |||  | |     |||   | | |||||||    |||    ||  || |  |    ||
Db      31393 TTTATATATATATAAAATATATATAAAATAAGATTTTATATATATATAAAATATATATAA 31452

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
              || |   || ||  ||| || ||  | |      |  || |  || |  ||  |
Db      31453 AATAAGATTTTATATATATATAAAATATATATAAAATAATATATAAAATAAGATTTCATA 31512

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAG-- 1171
              | |  |||||   || |    | |  |    |   |   || || ||| |  ||
Db      31513 TATATAAAATATATAAAATATATATAAAATATGATTTTTTATATATAAAAATATATATAA 31572

Qy       1172 ---AAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAAT 1228
                |||||||  ||| | ||||   | ||  || ||     | | ||    ||| | | |
Db      31573 AATAAGATTTTATATATAAAATATATAAAATATATATAAAATAAGATTTTATATATATAA 31632

Qy       1229 CATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGG 1288
              ||  ||     ||| |  | || | ||| ||||   | | |||      |       |
Db      31633 AATATGATTTTATATATATAAAATATATATAAAATATATATAAAATAAGATTTTATATAT 31692

Qy       1289 ATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348
              |||| |  | || |  ||   |||||     | | ||    |||| ||| :  |  ||
Db      31693 ATAAAATATATATAAAAATATATAAAATAAGATTTTATATATATATGAAATRTAAATAAG 31752

Qy       1349 ATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTG 1408
              ||  ||   |  |    || | |||  | |  ||     | | |     | ||| | |
Db      31753 ATTTTATATATATATGAAATGTAAATAAGATTTTATATATTTGAAATATAAATAAAATAA 31812

Qy       1409 ACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATT 1468
                | |     | || |   | ||      | |  ||| ||||      |  |||||
Db      31813 GATTATATCTAAATATATATAAAATAAGATTTTACATATATGAAATATATATAAAATAAG 31872

Qy       1469 TTATGA 1474
                | | |
Db      31873 GTTTTA 31878


RESULT 16
US-10-995-561-13303
; Sequence 13303, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13303
;   LENGTH: 201990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(201990)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13303

  Query Match             4.3%;  Score 63.6;  DB 6;  Length 201990;
  Best Local Similarity   45.1%;  Pred. No. 0.43;
  Matches  237;  Conservative    0;  Mismatches  289;  Indels    0;  Gaps    0;

Qy        879 AGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACC 938
              || |  | |  ||||| | ||   | ||  |  | | || | | ||  |  |  || |
Db      78483 AGAAGCAGGAAATATAAAAAATTTATATAATATATATAAAATATATTATATATAAATATA 78542

Qy        939 AGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATAT 998
              || | | | ||   ||      | | ||| | ||  ||| || |       |||| ||
Db      78543 TAAATATATAAAAATAAATATATATTATATATATAAAAATATATAAATATATATAAAAT 78602

Qy        999 TTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGA 1058
              | | |   || |   | | | |||   ||         ||  ||||  |    |  ||
Db      78603 ATATAAATATATTATATATAATATATTAAATATATTATATATAATATATTTATATAATAT 78662

Qy       1059 CGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGT 1118
              ||    || || || || |   | ||    || | |       ||| ||    |||
Db      78663 ATATAATATATTATATAAATATATAAATATATAAAAATATATAAAATATATAAATATATA 78722

Qy       1119 GCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTT 1178
              |     || |  |     | | |  |||   |  |||| | || || || || | |
Db      78723 ATATATATTATATATTAATATATTATATATTATATATAAATATATAAAAATATAATATAT 78782

Qy       1179 GATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGA 1238
              | | | | || |||| ||    || ||     |  ||    | |  || | | |
Db      78783 ATAAAATATATATAAAAATATATGATATATAATATATATAAATATATATGTATAAAATAT 78842

Qy       1239 ATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTT 1298
              ||  | |  |  | | |  | | |    |  |||||  ||||| |   |||| | |
Db      78843 ATATTATATATTATATACTATATATTATATATAATATATAAAAAATATATATAAAATATA 78902

Qy       1299 TAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCA 1358
              || |  |  | |   ||  |   | |  |||   |   | ||   || | ||  ||   |
Db      78903 TATATAACTTCACGGAAACTCTGAAGGAGAGGTCAAATATAATATTATATATTTTATATA 78962

Qy       1359 GGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
                | | |  |  | |   |||| |    | ||   | ||| | | |
Db      78963 TATAATTTATATATATATAAAACTATATATATATATTTTTGAGAGA 79008


RESULT 17
US-11-082-746-2/c
; Sequence 2, Application US/11082746
; Publication No. US20050244861A1
; GENERAL INFORMATION:
;  APPLICANT: Sundstrom, Paula
;  TITLE OF INVENTION: NUCLEIC ACIDS ESSENTIAL FOR EXPRESSION OF HYPHAL-SPECIFIC GENES
;  TITLE OF INVENTION:  AND METHODS FOR USING THE SAME
;  FILE REFERENCE: DC-0300
;  CURRENT APPLICATION NUMBER: US/11/082,746
;  CURRENT FILING DATE:  2005-03-17
;  PRIOR APPLICATION NUMBER: US 09/725,010
;  PRIOR FILING DATE: 2000-11-29
;  PRIOR APPLICATION NUMBER: US 60/167,672
```

```
;  PRIOR FILING DATE: 1999-11-29
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
;   LENGTH: 1434
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-11-082-746-2

  Query Match             4.3%;  Score 63.2;  DB 9;  Length 1434;
  Best Local Similarity   52.6%;  Pred. No. 0.48;
  Matches  160;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy       1107 TAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGA 1166
              | |  | | ||   ||  | | || |  | |    || |||| |  || |||||| || |
Db        305 TTAAACCACTGAATAAGCTTAGAACTAACTATTGTAAAAAAAAAGGGGTATTTATTAAAA 246

Qy       1167 ACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAA 1226
              |||  |  | |||     ||| | |||      | |    ||| |||| || ||| ||
Db        245 GCTATCAACTATGAACCGAAAACAGAAAACAAAGCAAGTAGAAGAAGAAGAAAAAGATAA 186

Qy       1227 ATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAA 1286
              |  ||  ||| ||     |    ||||| ||||| |  |  | | | |   |  |||
Db        185 ACAATTTGAACAAAACAATTTTTAAAAACTAAAAGTACCTATCACCTTTAATGTAGTAAA 126

Qy       1287 GGATAATAATTTTAAACCAAATGATAAAAGTTTGT-ATGATGAGCATATTAAAAAATATA 1345
              | |  ||| || ||  ||||||    ||  | ||  |||  |    |||    | |
Db        125 CAAAACAAATATTCAATAAAATGACTCTTGTCAATAATATTGAAAACCCAAAATTGAAAA 66

Qy       1346 AAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATAT 1405
              | ||| | ||  ||  || ||| ||    |||||   | ||| | | || |||  |||
Db          65 ACAATCAAAATTAGAAAAAAAAGAAATCCCAAAAAAAGAACAAAACTTAGTATCAGTTAT 6

Qy       1406 TTGA 1409
              | ||
Db           5 TAGA 2


RESULT 18
US-10-909-125-825
; Sequence 825, Application US/10909125
; Publication No. US20050261218A1
; GENERAL INFORMATION:
;  APPLICANT: Esau, Christine
;  APPLICANT:  Lollo, Bridget
;  APPLICANT:  Bennett, C. Frank
;  APPLICANT:  Freier, Susan M.
;  APPLICANT:  Griffey, Richard H.
;  APPLICANT:  Baker, Brenda F.
;  APPLICANT:  Vickers, Timothy
;  APPLICANT:  Marcusson, Eric G.
;  APPLICANT:  Koller, Erich
;  APPLICANT:  Swayze, Eric
;  APPLICANT:  Jain, Ravi
;  APPLICANT:  Bhat, Balkrishen
;  APPLICANT:  Peralta, Eigen
;  TITLE OF INVENTION: Oligomeric Compounds And Compositions For Use In Modulation
;  TITLE OF INVENTION:  Of Small Non-Coding RNAs
;  FILE REFERENCE: ISIS0080-100 (CORE0016US)
;  CURRENT APPLICATION NUMBER: US/10/909,125
;  CURRENT FILING DATE:  2004-07-30
;  PRIOR APPLICATION NUMBER: US 60/492,056
;  PRIOR FILING DATE: 2003-07-31
;  PRIOR APPLICATION NUMBER: US 60/516,303
;  PRIOR FILING DATE: 2003-10-31
;  PRIOR APPLICATION NUMBER: US 60/531,596
;  PRIOR FILING DATE: 2003-12-19
;  PRIOR APPLICATION NUMBER: US 60/562,417
;  PRIOR FILING DATE: 2004-04-14
;  NUMBER OF SEQ ID NOS: 2184
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 825
;   LENGTH: 6189
;   TYPE: DNA
;   ORGANISM: H. sapiens
US-10-909-125-825

  Query Match             4.3%;  Score 63.2;  DB 6;  Length 6189;
  Best Local Similarity   42.2%;  Pred. No. 0.48;
  Matches  517;  Conservative    0;  Mismatches  688;  Indels   21;  Gaps    2;

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              | | |||  ||   | | |    |   || |||  |||  ||||   | | |
Db       1894 ACATTTGAAAATTCAGATTGCACACTGAAAAAAATGGATAAAGAAGGTAAAACATTAAAA 1953

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||| | || ||| |  |  || | || ||    |||| |  | | |  |  | |||
Db       1954 AAACATAAATTGAAGCATAAAGAGAGGGAAAAAGAAAAGCATAAAAAAGAAATTGAAGGT 2013

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              | ||| || | |   |  ||   | | |    |   |  |  ||
Db       2014 GAAAAGGAAAAATACAAAACTAAGGATAGTGCCAAAGAACTGCAGAGGAGTGTGGAATTT 2073

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              |   |    ||  || | |     |  | ||  |    |||| |
Db       2074 GATAGAGAATTTTGGAAAGAGAATTTTTTTAAAAGTGATGAAACTGAAGATCTCTTTTTA 2133

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              || | ||||| ||       |  ||||||      | ||   |||  |  |  ||
Db       2134 AATATGGAACATGAATCCTTAACATTAGAAAAAAAATCAAAATTGGAAAAAAACATCAAA 2193

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              | || | |    | || ||||||      | || | ||  | ||  || | ||  ||
Db       2194 GATGATAAATCAACCAAGGAAAAGCATGTGTCAAAAGAGAGGAACTTTAAAGAGGAACGA 2253

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              |  |||||     ||  ||          ||| ||  | |  |  |||
Db       2254 GACAAGATTAAAAAGGAAAGCGAGAAATCTTTTAGGGAGGAAAAAATAAAAGATCTAAAA 2313

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||| ||   || |  ||  |  |||  |||   |||     | ||  |  | |
Db       2314 GAAGAGAGAGAAAACATACCCACAGATAAAGACTCAGAATTTACTTCTTTGGGTATGAGT 2373

Qy        613 CAAGAGCAGCAA---------AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAT 663
                 || ||             ||||||  |||  | |    || |  ||||| |
Db       2374 GCCATTGAGGAATCTATAGGGCTTCATTTAGTGGAAAAGGAAATAGACATTGAAAAACAA 2433

Qy        664 TTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA 723
              | |   | |||||||  |  ||| | |   |  ||  |  |  |    | ||
Db       2434 GAAAAGCATATAAAGGAAAGTAAAGAAAAACCTGAGAAGCGATCTCAAATTAAAGAAAAG 2493

Qy        724 ATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCT 783
              |  |     |  |   |  ||||  | | |  | |||  |||| |   |
Db       2494 GACATTGAGAAGATGGAAAGAAAAACCTTTGAAAAAGAAAAGAAGATAAAACATGAGCAT 2553

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA 843
              | |   | |||||   ||   ||||  |  |  |    | |||||      ||  ||
Db       2554 AAGTCAGAAAAAGACAAATTAGATCTTAGTGAATGTGTTGATAAAATAAAAGAAAAGGAC 2613

Qy        844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903
              ||  | ||    |   |   || |    | ||| | |   | | |   | ||
Db       2614 AAGCTATATTCGCATCACACAGAAAAATGCCATAAAGAAGGTGAGAAGAGCAAAAATACT 2673

Qy        904 CATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
              |  |    |||| |  || |    |  |||     | ||||    ||||       |
Db       2674 GCTGCTATTAAAAAAACTGACGACAGAGAGAAAAGTAGAGAAAAGATGGATAGGAAACAT 2733

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
              |||| | |||| | |  |||||          || |   ||    ||  |
Db       2734 GACAAAGAAAAGCCTGAAAAAGAGA-----------GGCATCTAGCAGAAAGCAAAGAA 2781

Qy       1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAA 1083
              |  | |    |||  ||| || |    |||  ||  ||      |     | ||  ||||
Db       2782 AAGCACTTGATGGAGAAAAAAAATAAACAATCAGATAATAGTGAATACAGTAAATCAGAA 2841

Qy       1084 GATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
```

```
Db       2842 AAAGGCAAAAATAAAGAAAAAGACAGGGAGCTAGATAAAAAGGAAAAATCTAGAGATAAA 2901

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203

Db       2902 GAAAGTATAAATATAACTAACTCCAAACACATACAGGAAGAAAAAAAATCAAGTATAGTA 2961

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263

Db       2962 GACGGTAATAAAGCACAACATGAAAAACCCTTATCCCTTAAAGAAAAAACAAAAGATGAA 3021

Qy       1264 GGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTAT 1323

Db       3022 CCTTTGAAAACTCCAGATGGAAAAGAAAAAGATAAAAAAGATAAAGATATAGATAGATAC 3081

Qy       1324 GATGAGCATATTAAAAAATATAAAAA 1349

Db       3082 AAAGAACGAGACAAACATAAAGATAA 3107


RESULT 19
US-10-995-561-13236
; Sequence 13236, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13236
;  LENGTH: 305312
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(305312)
;  OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13236

  Query Match             4.3%;  Score 63;  DB 6;  Length 305312;
  Best Local Similarity   51.6%;  Pred. No. 0.52;
  Matches  144;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy       1155 AATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGA 1214

Db     140661 AATAAAATAAATAACATAACATAACATAACATAACATAACATAACATATAAAATA 140720

Qy       1215 AGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATA 1274

Db     140721 AAATAAAATATAAAATAAAATAAAATAAAATAAAATAAAAATAAAATAAAATATAAAATA 140780

Qy       1275 TGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATAT 1334

Db     140781 AAATAAAATAAAATAAAATAAATAAAATAAAATAAAATAAAATATAAAATAAAATAAAAT 140840

Qy       1335 TAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATT 1394

Db     140841 AAAATAAAATAAAAATAAAATAAAATAAAATAAAATAAAATATAAAATAAAATAAATAAA 140900

Qy       1395 GTTTCATATATTTGACGGAGACAATGAAATTTTACAGAT 1433

Db     140901 ATAAAATAAAATAAAATAAAATAAAATAAAAAAATAAAT 140939


RESULT 20
US-10-793-626-693
; Sequence 693, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 693
;  LENGTH: 627
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: synthetic
;  OTHER INFORMATION: nucleic acid sequence
US-10-793-626-693

  Query Match             4.2%;  Score 62.8;  DB 6;  Length 627;
  Best Local Similarity   48.6%;  Pred. No. 0.54;
  Matches  203;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy       1061 ATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGC 1120

Db        134 ATCAAGTAAGTAAAAAAATAAACGAATTAGAAGAGAAAAAGCAAAAATTATTTAAAAAGG 193

Qy       1121 AATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGA 1180

Db        194 TAAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAA 253

Qy       1181 TAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAAT 1240

Db        254 AACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCAT 313

Qy       1241 TATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA 1300

Db        314 TTAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTG 373

Qy       1301 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGG 1360

Db        374 AACAACTTGATAGTGCTAT---TAAAGAAAAATATAAATCACATGATGCTTATGCAAAGG 430

Qy       1361 TTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATG 1420

Db        431 CTTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATG 490

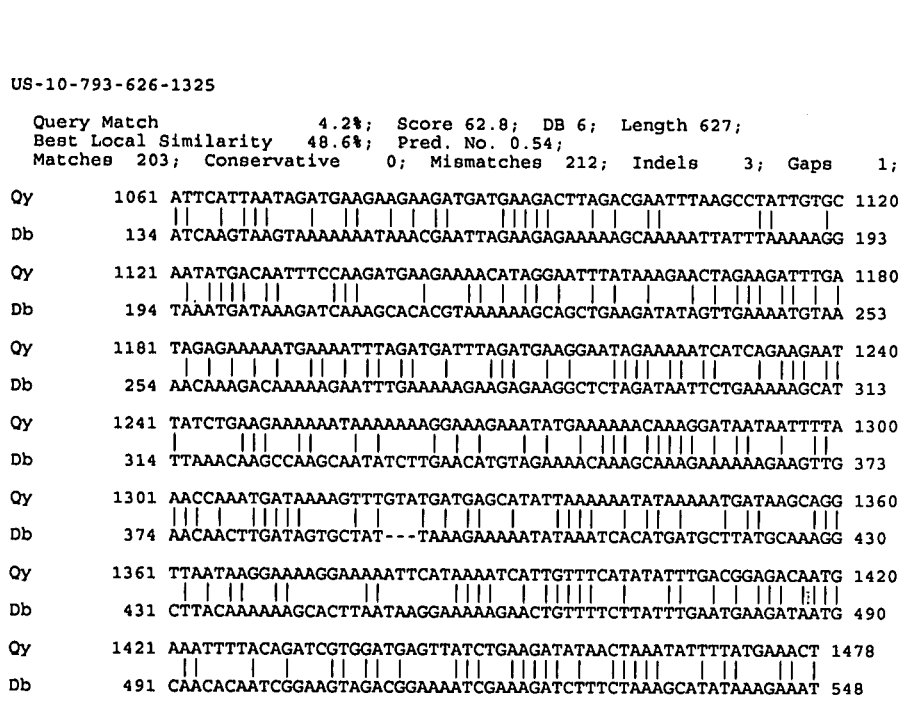
Qy       1421 AAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db        491 CAACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 548


RESULT 21
US-10-793-626-1325
; Sequence 1325, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1325
;  LENGTH: 627
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence: synthetic
;  OTHER INFORMATION: nucleic acid sequence
```

US-10-793-626-1325

Query Match 4.2%; Score 62.8; DB 6; Length 627;
Best Local Similarity 48.6%; Pred. No. 0.54;
Matches 203; Conservative 0; Mismatches 212; Indels 3; Gaps 1;



RESULT 22
US-10-793-626-4039/c
; Sequence 4039, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4039
; LENGTH: 2999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4039

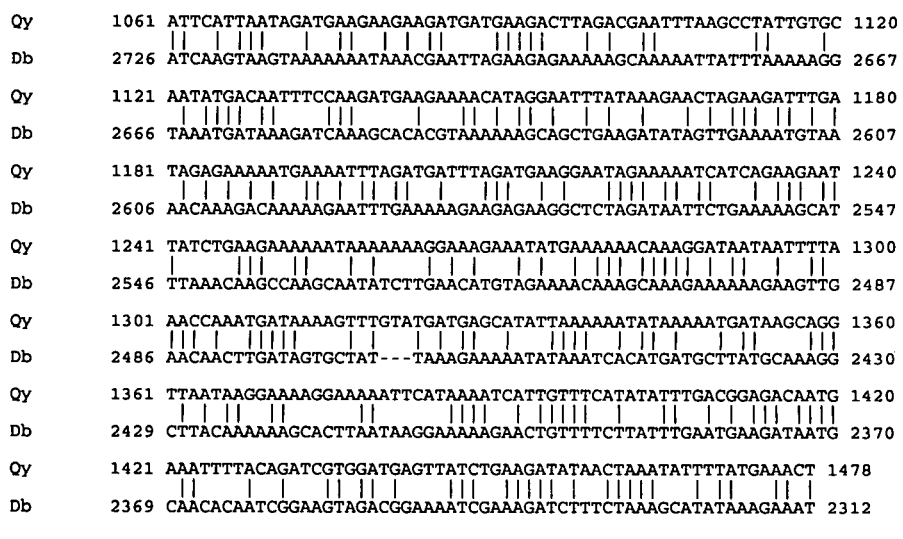
Query Match 4.2%; Score 62.8; DB 6; Length 2999;
Best Local Similarity 48.6%; Pred. No. 0.54;
Matches 203; Conservative 0; Mismatches 212; Indels 3; Gaps 1;





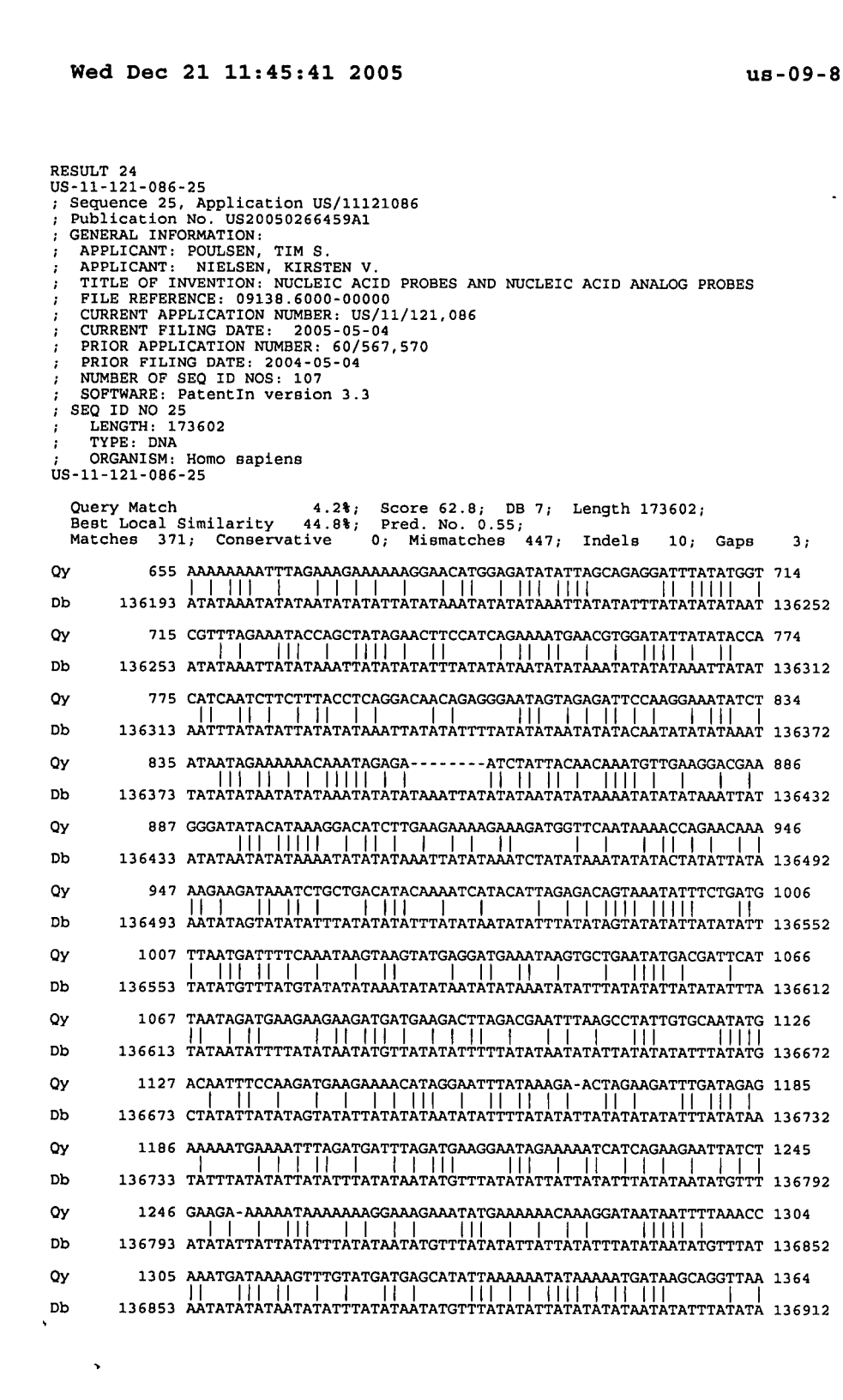
RESULT 23
US-10-793-626-3338/c
; Sequence 3338, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3338
; LENGTH: 3691
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3338

Query Match 4.2%; Score 62.8; DB 6; Length 3691;
Best Local Similarity 48.6%; Pred. No. 0.54;
Matches 203; Conservative 0; Mismatches 212; Indels 3; Gaps 1;



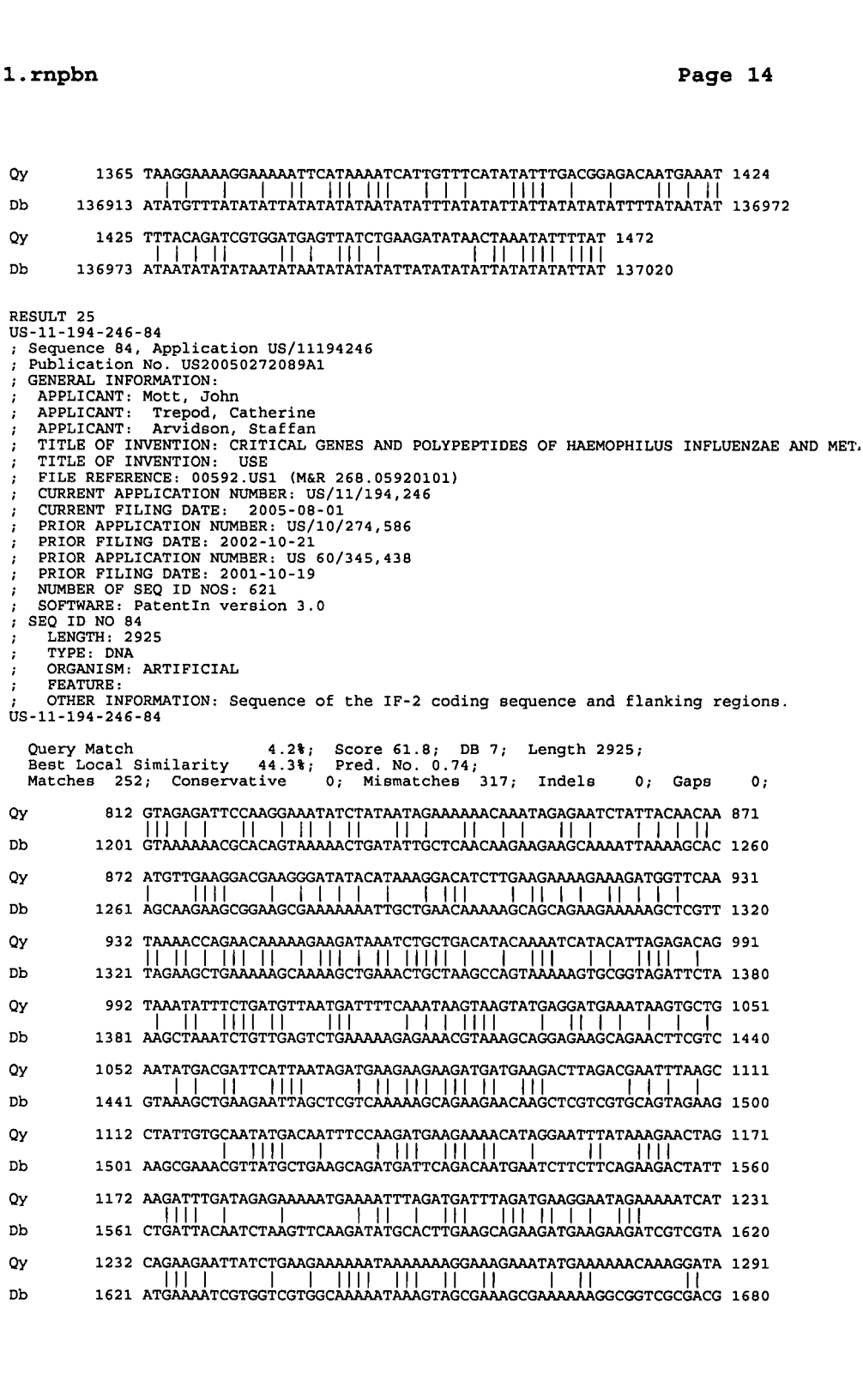
RESULT 24
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25

Query Match 4.2%; Score 62.8; DB 7; Length 173602;
Best Local Similarity 44.8%; Pred. No. 0.55;
Matches 371; Conservative 0; Mismatches 447; Indels 10; Gaps 3;



RESULT 25
US-11-194-246-84
; Sequence 84, Application US/11194246
; Publication No. US20050272089A1
; GENERAL INFORMATION:
; APPLICANT: Mott, John
; APPLICANT: Trepod, Catherine
; APPLICANT: Arvidson, Staffan
; TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND MET
; TITLE OF INVENTION: USE
; FILE REFERENCE: 00592.US1 (M&R 268.05920101)
; CURRENT APPLICATION NUMBER: US/11/194,246
; CURRENT FILING DATE: 2005-08-01
; PRIOR APPLICATION NUMBER: US/10/274,586
; PRIOR FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US 60/345,438
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 621
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84
; LENGTH: 2925
; TYPE: DNA
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: Sequence of the IF-2 coding sequence and flanking regions.
US-11-194-246-84

Query Match 4.2%; Score 61.8; DB 7; Length 2925;
Best Local Similarity 44.3%; Pred. No. 0.74;
Matches 252; Conservative 0; Mismatches 317; Indels 0; Gaps 0;



```
Qy       1292 ATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATG 1351
Db       1681 ATGAAAATAGCAAAAACTCGAAAAACGAGCGCGAATCTAATCGTAAAAATCAAAAAGACG 1740

Qy       1352 ATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1741 CTAAGTTTGGTAAAGGTAAAAATGGCAAA 1769


RESULT 26
US-10-995-561-13443/c
; Sequence 13443, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13443
;   LENGTH: 49979
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(49979)•
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13443

  Query Match             4.2%;  Score 61.8;  DB 6;  Length 49979;
  Best Local Similarity   48.4%;  Pred. No. 0.75;
  Matches  220;  Conservative    3;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
Db      29303 GGCCTAGAAACTCTTTTATTTTAAAAAATATATGTATATTATATTTATATATATATAATA 29244

Qy       1024 AGTAAGTATGAGGATGAAATAAGT-GCTGAATATGACGATTCATTAATAGATGAAGAAGA 1082
Db      29243 TATAAAAATGTATATATAATATGTATATAAATATAATGTATAAATATAATATTTATATAT 29184

Qy       1083 AGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGA 1142
Db      29183 ATTTATAAATATATTTATATATTATATTTATAATATATATTATATATATAATATATATAA 29124

Qy       1143 AGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGA 1202
Db      29123 TAAATATATGTACAATTATATATWWTATATATAATATATATAATATATATTATAGATATA 29064

Qy       1203 TGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAA 1262
Db      29063 TTAAATATATATATTTTTATATATATTATATATTTATTATATATATTATATATAAATAAT 29004

Qy       1263 AGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTA 1322
Db      29003 ATAAATATAATATGTATTATATATAAATAATATAAATATAATATATATAAMCAATATAAA 28944

Qy       1323 TGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATT 1382
Db      28943 TATATTACATA-TAAATAATATAATATATATAATATATTATATAAATAATATAAATATAA 28885

Qy       1383 CATAAAATCATTGTTTCATATATTTGACGGAGACA 1417
Db      28884 TATATATTACATATGTAATATAAATAATATAGATA 28850


RESULT 27
US-10-995-561-13308
; Sequence 13308, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13308
;   LENGTH: 148935
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13308

  Query Match             4.2%;  Score 61.6;  DB 6;  Length 148935;
  Best Local Similarity   43.3%;  Pred. No. 0.8;
  Matches  257;  Conservative    6;  Mismatches  331;  Indels    0;  Gaps    0;

Qy          27 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACT 86
Db       22401 ACAGAGGCGTCCTAGTCACACTGCTAAAGATCACAGTCTTAAAATCCATGCACAGAGGCR 22460

Qy          87 TGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGA 146
Db       22461 TCYTAGTCACACTGCTRAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTGGTCA 22520

Qy         147 AAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 206
Db       22521 CACTGCTAAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTRGTCACACTGCTGA 22580

Qy         207 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 266
Db       22581 AGATCACARTCTTAAAATCCATGCACAGAGGCRYCCTAGTCACACTGCTGAAGATCACCV 22640

Qy         267 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 326
Db       22641 TCTTAAAATCCATGCACAGAGGCACCCTAGTCACACTGCTGAAGATCACCATCTTAAAAT 22700

Qy         327 AGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGA 386
Db       22701 CCATGCACAGAGGCGTCCTAGTCACACTGCTGAAGATCACCATCTTAAAATCCGTGCACA 22760

Qy         387 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGC 446
Db       22761 GAGGCGTCCTAGTCACACTGCTGAAGATCACAGTCTTAAAATCCATGCACAGAGGCACCC 22820

Qy         447 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAA 506
Db       22821 TAGTCACACTGCTGAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTGGTCACAC 22880

Qy         507 GTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 566
Db       22881 TGCTGAAGATCACAGTCTTAAAATCCATGCACAGAGGCACCCTAGTCACACTGCTGAAGA 22940

Qy         567 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCA 620
Db       22941 TCACCATCTTAAAATCCATGCACAGAGGCGTCCTAGTCACACTGCTAAAGATCA 22994


RESULT 28
US-11-121-086-60/c
; Sequence 60, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
```

```
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
;   LENGTH: 191091
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-60

  Query Match             4.1%;  Score 60.6;  DB 7;  Length 191091;
  Best Local Similarity   49.3%;  Pred. No. 1.1;
  Matches  185;  Conservative    0;  Mismatches  189;  Indels    1;  Gaps    1;

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              ||| | | || |     ||||||||  | | || | | | ||  | | |   | |    |
Db      40243 GAAAGAAAAGGAAGAAGGGAAGAAAAGGAGAGAAAAAGAGAAGGAAAGAAAGGAGAGG 40184

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
                | | |  |||||||||     ||| | |  || | ||   |||| | |   |
Db      40183 GAAAGAGGGAAAGAAAAGAAAGAAG-AGAGAAAGGAGAGAAAGAAAAGAAAGAAGAGAGA 40125

Qy        448 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 507
              || ||||    | ||  | |  |||  ||   |    | |||  |          ||| ||||
Db      40124 AAGGAAATAAAGAAAAGATAGAAAAAAGAAGGAAGGAAGGAGGAAAAAAAGGAAGGAAAG 40065

Qy        508 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 567
                ||||   | | || ||   | |  |||| |||      ||  || |   | |||||
Db      40064 AGAGAAGAGAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAAGAAAGAA 40005

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
              || ||||      ||||  | ||| | |     | | | ||       | || |  || ||
Db      40004 AAAGAAAGAAGGAAAGAAAGAAAGAAAAGAAAAAGAAAGAAAGAAGGAAAGAAAGAAGGA 39945

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
              |   ||||  || ||| |      | |||| ||        ||  || |||| || |  ||
Db      39944 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAGGAAAGAAAGAAAGAAAGAAAGAAGA 39885

Qy        688 GATATATTAGCAGAG 702
                | | |  |  ||||
Db      39884 AAGAAAGAAAGAGAG 39870


RESULT 29
US-10-995-561-13460
; Sequence 13460, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13460
;   LENGTH: 20317
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13460

  Query Match             4.0%;  Score 59.8;  DB 6;  Length 20317;
  Best Local Similarity   45.0%;  Pred. No. 1.4;
  Matches  284;  Conservative    5;  Mismatches  337;  Indels    5;  Gaps    2;

Qy        828 AATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAG 887
              |||||  ||| ||| | || | | ||      ||      | | | |  ||  |      |
Db      17975 AATATATATCATATATAATATATATTATATATTATACAATATATACATTATATATTATAT 18034

Qy        888 GGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAA 947
                    ||| |    | |  |  | |    | ||  |  |    |  | |  | |
```

```
Db      18035 ATTATACAATATATACATTATATATYATATACTATATATTATATATCATATAYAYTATAT 18094

Qy        948 AGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGT 1007
              | || ||| :     ||| |||| | || ||| || | | |   || |   | ||    |
Db      18095 ATAATATAYWATATATGATATACTATATAATATATAATATATGATATAATATACTATGAT 18154

Qy       1008 TAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATT 1067
                || || |   |   |  | ||| |  |  | ||||     ||      |  ||  | |
Db      18155 ATATAATATATGATATATGATATATCA--AATATATAATATACGATATATAATATATAAT 18212

Qy       1068 AATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGA 1127
              |   |||      | |  ||     | :  ||     |  |||    |  | |   ||||
Db      18213 ATATGATATCATATATTATATATTATRTATATTATATATTATATATTATATTTTATATAT 18272

Qy       1128 CAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAA 1187
              | |   |      | | | ||||   | |    | ||  ||  | || | ||| | ||
Db      18273 KATATATGATATATCATATATCATATATAATATATATGATATATSATATATCATATATAA 18332

Qy       1188 AAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGA 1247
              |   | | |||| ||  | || ||||   |||| | | |  ||  ||    || || |
Db      18333 TATATATATTTTATATATTATATATGATATAATATATATAATATATGATACATGATATAT 18392

Qy       1248 AGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAA 1307
              |  |  |   |  | |  | | |       | || | | |  ||||||  | |||   | |
Db      18393 AATATTATATATTATATAATATATTACATATAATATATAATATAATATATATAATATATA 18452

Qy       1308 TGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAA 1367
                ||| ||  |  |||       ||| || | |  ||||  ||  ||  ||  ||
Db      18453 CTATATAATATAATATAGGATATATAATATATATAATAATTATATTATATAGTATATATT 18512

Qy       1368 GGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTT 1427
              | |     | |:|  |||  ||  :   |   |||||   |   | | ||| |||||  :
Db      18513 ATATATATTATAWTATATACTATATWATATAATTATAT---AATTATATAATTAAATTAW 18569

Qy       1428 ACAGATCGTGGATGAGTTATCTGAAGATATA 1458
              | |  |     |||| |||| | |  |||||
Db      18570 ATAATTAAATTATGATTTATATAATTATATA 18600


RESULT 30
US-10-995-561-13274/c
; Sequence 13274, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13274
;   LENGTH: 415117
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(415117)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13274

  Query Match             4.0%;  Score 59.6;  DB 6;  Length 415117;
  Best Local Similarity   46.8%;  Pred. No. 1.5;
  Matches  259;  Conservative    1;  Mismatches  285;  Indels    9;  Gaps    2;

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              ||| |||||     |||   ||||||    | ||  | |  |      |  ||||  |
Db      69341 AGASCAAGACCCTGTTGAAGAAGAAAGAAGGAAAAGAAAGGAGGAAAAGAAAGAAAGAAG 69282

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
                | | |    || |||  |  |  | || |    |  ||     || | || |||   |
```

```
Db      69281 GGAAAGAAGGAAGGAAGGGAGGGGAAAAAAGAGGAAAGAAAAGAAAGAGGAAAGAAAGGA 69222

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302

Db      69221 AAGGAAAGGGAAAAGAGAAGGAAAAAG-AAAAGGGAAGGAAGGAAAGAAGGGGAAGGAAA 69163

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362

Db      69162 GGAAARAGAAAAGAGAAGGAAAATGAAAAGAGAAGGAAAAAGAAAAGGAAGGAAGGAAGG 69103

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422

Db      69102 AAAAAAGAGGAAGAAAAGGAAAGAGA--------AAAGAGAAGGAAAAAGAAAAGAGAAG 69051

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482

Db      69050 GAAGGAAGGAAAAGAGAGAGGAGAGAGGAAAGGAAGAAAAGGGAAAAGGGGAAGGTGAGG 68991

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542

Db      68990 AGAGGAGAGACGAGAAAGGAGAAAGAAGAAAGGAGGGGAAAGGGGAATGGGAAAAAGGGA 68931

Qy        543 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 602

Db      68930 AAGGGGAAAGGAAAGGGAAAAGAAGAAAAGAAGGAGGGAGGGAGGAAGGGGAGGAGGAGG 68871

Qy        603 AGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAA 662

Db      68870 AGAGGAAGAGAGAGAAGGAGGAGGAGAGGAACAGGAGGAGGAGAAGAAGAAGGAAGGAAG 68811

Qy        663 TTTAGAAAGAAAAA 676

Db      68810 AAGGAAAAGAAGAA 68797

Search completed: December 20, 2005, 23:15:48
Job time : 422.691 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:31:02 ; Search time 55.3022 Seconds
(without alignments)
5156.873 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093216_22943/app_query.fasta_1.3740
-DB=PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_179_@runat_21122005_093216_22943 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2406.5	96.8	1909	2	A45592	liver stage antige
2	380.5	15.3	1407	1	S28589	trichohyalin - rab
3	380.5	15.3	1549	1	A40691	trichohyalin - she
4	376	15.1	1898	1	A45973	trichohyalin - hum
5	356	14.3	1017	2	T15598	hypothetical prote
6	353.5	14.2	737	2	T15597	hypothetical prote
7	332.5	13.4	1192	2	A71623	probable secreted
8	326.5	13.1	1979	1	S03166	myosin heavy chain
9	323.5	13.0	1039	2	S18199	myosin heavy chain
10	321	12.9	771	1	A33430	h-caldesmon - chic
11	319	12.8	1938	2	JC5421	smooth muscle myos
12	319	12.8	1972	2	JC5420	smooth muscle myos
13	318.5	12.8	1961	1	A61231	myosin heavy chain
14	317.5	12.8	1166	2	H86341	hypothetical prote
15	317.5	12.8	1937	2	I38055	myosin heavy chain
16	317.5	12.8	1992	2	A47297	myosin heavy chain
17	317	12.8	1432	2	B85431	trichohyalin like
18	316	12.7	1940	2	A29320	myosin heavy chain
19	314	12.6	1938	2	A59293	skeletal myosin he
20	314	12.6	1976	2	A59252	myosin heavy chain
21	313.5	12.6	1738	2	T14867	interaptin - slime
22	312.5	12.6	1959	1	A33977	myosin heavy chain
23	312	12.6	2007	1	B43402	myosin heavy chain
24	311.5	12.5	2139	2	T18296	myosin heavy chain
25	310.5	12.5	665	2	B71609	hypothetical prote
26	310.5	12.5	678	2	A54514	glutamic acid-rich
27	310	12.5	1963	1	MWKW	myosin heavy chain
28	310	12.5	2116	2	A26655	myosin heavy chain
29	309.5	12.4	1956	2	T16416	hypothetical prote
30	309	12.4	1931	2	A59234	slow myosin heavy
31	309	12.4	1972	1	A41604	myosin heavy chain
32	307.5	12.4	1701	2	T09127	probable erythrocy
33	307	12.3	746	2	T47237	myosin II heavy ch
34	306.5	12.3	522	2	C96608	hypothetical prote
35	306.5	12.3	1790	2	S67593	transport protein
36	306	12.3	1999	1	S21801	myosin heavy chain
37	305.5	12.3	2017	1	A36014	myosin heavy chain
38	305.5	12.3	2057	2	S61477	myosin II heavy ch
39	305	12.3	1964	2	A59282	nonmuscle myosin I
40	304.5	12.2	1939	1	A46762	myosin alpha heavy
41	304	12.2	451	2	G70241	hypothetical prote
42	303.5	12.2	1938	2	I49464	alpha cardiac myos
43	303.5	12.2	1939	2	I48175	myosin heavy chain
44	303	12.2	1526	2	A45605	mature-parasite-in
45	302.5	12.2	1130	2	T34081	hypothetical prote

ALIGNMENTS

RESULT 1
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C;Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R;Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A;Reference number: S24597
A;Accession: S24597
A;Molecule type: DNA
A;Residues: 1-1909 <ZHU>
A;Cross-references: UNIPROT:Q25893; UNIPARC:UPI0000081753; EMBL:X56203; NID:g9915; PID:g
R;Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A;Title: Structure of Plasmodium falciparum liver stage antigen-1.
A;Reference number: A45592; MUID:92107224; PMID:1840628
A;Accession: A45592
A;Molecule type: DNA
A;Residues: 1-195;638-688;1165-1215;1590-1909 <ZH2>
A;Cross-references: UNIPARC:UPI0000177378; UNIPARC:UPI0000177379; UNIPARC:UPI000017737A;
A;Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A;Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A;Reference number: S29393; MUID:87315391; PMID:3306406
A;Accession: S29393
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 323-387 <GUE1>
A;Cross-references: UNIPARC:UPI000017737C; EMBL:M28266
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A;Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by
A;Reference number: S34842
A;Accession: S34842
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 323-381,'HKAI' <GUE2>
A;Cross-references: UNIPARC:UPI000017737D; EMBL:M28266
A;Note: difference at carboxyl end due to frameshift error
C;Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: EF hand
F;154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Alignment Scores:
Pred. No.:              2.57e-119      Length:         1909
Score:                  2406.50        Matches:        487
Percent Similarity:     89.89%         Conservative:   2
Best Local Similarity:  89.52%         Mismatches:     4
Query Match:            96.80%         Indels:         51
DB:                     2              Gaps:           2

US-09-837-344-41 (1-1482) x A45592 (1-1909)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT------------------------------- 87
              |||||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405

Qy         88 ------------------GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                              ||||||||||||||||||||||||||||||||||||||||||||
Db       1406 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 1425

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 1445

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 1465

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 309
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db       1466 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 1485

Qy        310 GAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 369
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1486 GluGlnGlnArgAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 1505

Qy        370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 1525

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
              ||||||||||||||||  ||||||||||||||||||  ||||||||||||||||
Db       1526 GluGlnGluArgLeuAlaAsnGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnGlu 1545

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1546 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1565

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------------ 597
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1585

Qy        597 ------------------------------------------------------------ 597

Db       1586 LeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlu 1605

Qy        598 ----------------------------GCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                                         ||||||||||||||||||||||||||||||
Db       1606 GlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArg 1625

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 AspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGly 1645

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1646 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 1665

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1666 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 1685

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1686 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 1705

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1706 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 1725

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1726 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 1745

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 1765

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1766 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 1785

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1786 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 1805

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 1825

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1826 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 1845

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1846 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 1865

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1866 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 1885

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1886 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 1905

Qy       1468 TTTATGAAACTA 1479
              ||||||||||||
Db       1906 PheMetLysLeu 1909

RESULT 2
S28589
trichohyalin - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: S28589
R;Fietz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A;Description: Examination of the gene encoding rabbit trichohyalin.
A;Reference number: S28589
A;Accession: S28589
A;Molecule type: DNA
A;Residues: 1-1407 <FIE>
A;Cross-references: UNIPROT:P37709; UNIPARC:UPI000013738C; EMBL:Z19092; NID:g1746; PIDN:
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
```

C;Genetics:
A;Introns: 46/3
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.:	3.32e-13	Length:	1407
Score:	380.50	Matches:	127
Percent Similarity:	45.84%	Conservative:	132
Best Local Similarity:	22.48%	Mismatches:	167
Query Match:	15.31%	Indels:	139
DB:	1	Gaps:	20

US-09-837-344-41 (1-1482) x S28589 (1-1407)

```
Qy          1 CAAGAACAACAAAGCGATCTA---------GAACAAGAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||:::   :::|||         |||::::::|||:::   :::|||:::
Db        140 GlnGluGlnGluArgGluLeuAlaGluGluGluGluGlnArgLysLysArgGluArgPhe 159

Qy         52 CAAGAA------------CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG 99
              ::::::            |||       |||   ||||||   ||||||   :::::::::
Db        160 GluGlnHisTyrSerArgGlnTyrArgAspLysGluGln---ArgLeuGlnArgGlnGlu 178

Qy        100 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 159
              |||:::|||:::::::::       |||:::|||:::|||   :::   |||   :::::: 
Db        179 LeuGluGluArgArgAla------GluGluGluGlnLeuArgArgArgLysGlyArgAsp 196

Qy        160 CAACAAAGCGATCTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAA 210
              :::      :::|||:::|||         ||||||   :::::::::|||:::
Db        197 AlaGluGluPheIleGluGluGluGlnLeuArgArgArgGluGlnGlnGluLeuLysArg 216

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG--------------TTG 255
              :::      :::   ||||||:::||||||   :::   :::                |||
Db        217 GluLeuArgGluGluGluGlnGlnArgArgGluArgArgGluGlnHisGluArgAlaLeu 236

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              ||||||::::::      |||   :::   |||      :::|||:::   :::||||||
Db        237 GlnGluGluGluGluGlnLeuLeuArgGlnArgArgTrpArgGluGluProArgGluGln 256

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
              |||          |||   ||||||      |||::::::::::|||::::::
Db        257 Gln----------GlnLeuArgArgGluLeuGluGluIleArgGluArgGluGlnArg 272

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
              ||||||||||||   |||   :::::::::|||:::   :::|||      |||||||||
Db        273 LeuGluGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnArg---LeuGluGln 291

Qy        436 GAG--------------------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              |||                   |||         |||::::::::::|||::::::
Db        292 GluGluArgArgGluGlnGlnLeuArgArgGluLeuGluGluIleArgGluArgGluGln 311

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
              ||||||||||||   |||   :::::::::|||::::::::::::      :::
Db        312 ArgLeuGluGlnGluGluArgArgGluGlnArgLeuGluGlnGluGluArgArgGluGln 331

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 594
              |||   :::|||      |||:::::::::|||::::::      ||||||||||||
Db        332 GlnLeuLysArgGluLeuGluGluIleArgGluArgGluGlnArgLeuGluGlnGluGlu 351

Qy        595 CGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||   ::::::   |||   |||:::   |||
Db        352 ArgArgGluGlnLeuLeuAlaGluGluValArg-------------------------- 362

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                          |||:::   :::|||   |||:::   |||
Db        363 ------------GluGlnAlaArgGluArgGlyGluSerLeuThrArgArgTrpGlnArg 378

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
              :::||||||            |||       :::          :::   |||       |||
Db        379 GlnLeuGlu---------SerGluAlaGlyAlaArgGlnSerLysValTyrSerArgPro 395

Qy        775 ---------CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825
                                  ::::   ||||||   ||||||                |||      :::
Db        396 ArgArgGlnGluGluGlnSerLeuArgGlnAspGlnGluArgArgGlnArgGlnGluArg 415

Qy        826 GAAATATCTATAATAGAAAAAACAAATAGA---------------------------- 855
              |||       :::   |||:::      |||
Db        416 GluArgGluLeuGluGluGlnAlaArgArgGlnGlnGlnTrpGlnAlaGluGluGluSer 435

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT------ 909
              |||                 :::       |||       :::   :::       |||
Db        436 GluArgArgArgGlnArgLeuSerAlaArgProSerLeuArgGluArgGlnLeuArgAla 455

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAA------------CCAGAACAAAAAGAAGATAAA 957
              ||||||:::::::::         :::               ||||||:::   ::::::
Db        456 GluGluArgGlnGluGlnGluGlnArgPheArgGluGluGluGluGlnArgArgGluArg 475

Qy        958 TCTGCTGACATACAA---------------------------------------------- 972
                  ::::::|||
Db        476 ArgGlnGluLeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGln 495

Qy        972 ---------------------------------------------------------- 972

Db        496 GlnLeuGlnGluGluAspSerPheGlnGluAspArgGluArgArgArgArgGlnGlnGlu 515

Qy        973 ----------------------------------------------------AATCAT 978
                                                                     |||
Db        516 GlnArgProGlyGlnThrTrpArgTrpGlnLeuGlnGluGluAlaGlnArgArgArgHis 535

Qy        979 ACATTA--------------GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
              ||||||               :::   :::   :::   :::      :::   |||
Db        536 ThrLeuTyrAlaLysProGlyGlnGlnGluGlnLeuArgGluGluGluGluLeuGlnArg 555

Qy       1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATAT---GACGATTCATTAATAGATGAAGAA 1080
                 |||          |||      ||||||   ::::::      :::      |||
Db        556 GluLysArgArgGlnGluArgGluArgGluTyrArgGluGluGluLysLeuGlnArgGlu 575

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||::::::                                           :::
Db        576 GluAspGluLysArg---------------------------------------ArgArg 582

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::|||            |||:::|||||||||:::|||   ::::::   :::
Db        583 GlnGluArgGluArgGlnTyrArgGluLeuGluGluLeuArgGlnGluGluGlnLeuArg 602

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              |||                 |||:::      :::|||      ||||||:::::::::
Db        603 AspArgLysLeuArgGluGluGluGlnLeuLeuGlnGluArgGluGluGluArgLeuArg 622

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT------TTTAAACCAAATGATAAA 1314
              :::   ::::::      |||   ::::::               :::      ::::::
Db        623 ArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgGlnGluGluGln 642

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
                 |||      |||      :::|||   :::   ::::::|||:::   :::   :::
Db        643 GluLeuArgGlnGluArgGluArgLysLeuArgGluGluGluGluGlnLeuLeuArgArgGlu 662

Qy       1375 GAAAAATTCATAAAA 1389
              |||:::   ::::::
Db        663 GluGlnGluLeuArg 667
```

RESULT 3
A40691
trichohyalin - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: A40691; A34209; S32633
R;Fietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A;Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-bindi
A;Reference number: A40691; MUID:93260018; PMID:7684041
A;Accession: A40691
A;Molecule type: DNA
A;Residues: 1-1549 <FIE>
A;Cross-references: UNIPROT:P22793; UNIPARC:UPI000013738D; EMBL:Z18361; NID:g295940; PID

```
A;Note: sequence extracted from NCBI backbone (NCBIP:132511)
R;Fietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A;Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker
A;Reference number: A34209; MUID:90130632; PMID:2298812
A;Accession: A34209
A;Molecule type: mRNA
A;Residues: 1016-1151,1205-1257,1281-1398,'G',1400-1549 <FI2>
A;Cross-references: UNIPARC:UPI000016C5F4; GB:X51695; NID:g1827; PIDN:CAA35992.1; PID:g1
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Introns: 46/3
A;Note: single copy gene
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>
F;387-851/Region: 28-residue repeats
F;886-1519/Region: 23-residue repeats

Alignment Scores:
Pred. No.:                3.27e-13      Length:          1549
Score:                    380.50        Matches:         133
Percent Similarity:       43.59%        Conservative:    129
Best Local Similarity:    22.13%        Mismatches:      194
Query Match:              15.31%        Indels:          145
DB:                       1             Gaps:            18

US-09-837-344-41 (1-1482) x A40691 (1-1549)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::   |||:::      :::||||||:::::|||   :::   :::|||:::|||:::
Db        197 GluArgGlnArgLeuGlnValGluGlnGlnGlnArgGlnArgGlyGluLeuArgGluArg 216

Qy         61 CAAAGCGATTTA---------------GAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
              |||   ::::::                |||          |||   :::|||:::|||:::
Db        217 GlnGluGluValGlnLeuGlnLysArgGluThrGlnGluLeuGlnArgGluArgLeuGlu 236

Qy        106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
              |||:::|||         :::|||:::|||      :::|||:::|||   ||||||:::
Db        237 GluGluGlnGlnLeuGlnLysGlnLysArgGlyLeuGluGluArgLeuLeuGluGlnGlu 256

Qy        166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
                  ::::::   ||||||   :::::::::   :::|||||||||
Db        257 ArgArgGluGlnGluLeuArgArgLysGluGlnGluArgArgGluGlnGlnLeuArgGln 276

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------------------- 264
              |||||||||         :::|||::::::   |||:::
Db        277 GluGlnGluGluAlaThrGlnGluGluIleSerGluArgGlyGluSerArgThrSerArg 296

Qy        265 ---CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG------------------ 303
                 |||      ||||||   :::      |||:::::::::
Db        297 CysGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnArgLysValTyrSerArgPro 316

Qy        304 ---TTACAAGAGCAGCAAAGC-----------------------GATTTAGAACAAGAG 336
                 ||||||||||||||||                       :::   :::||||||
Db        317 HisArgGlnGluGlnGlnSerArgArgGlnGluGlnGluLeuLeuGluArgGlnGlnGlu 336

Qy        337 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------------TTAGAA 381
              ::::::      |||::::::|||      |||   |||                |||:::
Db        337 GlnGlnIleSerGluGluValGlnSerLeuGlnGluAspGlnGlyArgGlnArgLeuLys 356

Qy        382 CAAGAGAGACGT---------------------------------------------- 393
              ||||||:::|||
Db        357 GlnGluGlnArgTyrAspGlnAsnTrpArgTrpGlnLeuGluGluGluSerGlnArgArg 376

Qy        394 ---------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
                            ||||||         :::||||||      :::   ||||||
Db        377 ArgTyrThrLeuTyrAlaLysProAlaGlnArgGluGlnValArgGluGluGluGlnLeu 396

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA------ 492
              ||||||   :::||||||||||||   ::::::        |||:::||||||
Db        397 ArgLeuLysGluGluLysLeuGlnArgGluLysArgArgGlnGluArgGluArgGlnTyr 416

Qy        493 ------------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                               |||   :::|||:::||||||   ::::::      |||:::
Db        417 ArgGluValGluLeuGlnArgGluGluGluArgLeuGlnArgGluGluGluGlnLeuGln 436

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA-------------------------- 567
              :::|||   |||   |||   :::   ||||||
Db        437 ArgGluGluArgGluLysArgArgArgGlnGluArgGluLysGlnTyrLeuGluLysVal 456

Qy        568 ---------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
                       ::::::      |||:::::|||   |||   |||   :::   :::
Db        457 GluLeuTrpGluGluGluGlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArg 476

Qy        619 CAGCAAAGAGATTTAGAA----------------------------CAAAGGAAGGCT 648
              :::::::::   ||||||                            ||||||:::
Db        477 GluLysGlnTyrLeuGluLysValGluLeuArgGluGluGluGlnLeuGlnArgGlnGlu 496

Qy        649 GATACGAAAAAAAAATTTAGAAAGAAAAAAG---------GAACATGGAGATATATTAGCA 699
                   |||:::      ||||||::::::         |||      ::::::
Db        497 ArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGlu 516

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||:::          |||   |||             :::          |||   ||||||
Db        517 GluGluGlnLeuGlnArgGluGlu------ArgGluLysArgArgGlnGluArgGluArg 534

Qy        760 GGATATTAT------------ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
                 |||                :::      :::      |||   :::      |||
Db        535 GlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGlu 554

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAA--------- 858
                  |||      :::|||        :::||||||        :::
Db        555 LysArgArgGlnGluArgGluLysGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 574

Qy        859 ------TCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                          :::   |||::::::      |||   :::||||||
Db        575 GlnLeuGlnArgGlnGluArgGlnLysArgArgGlnGluArgGluLysGlnTyrLeuGlu 594

Qy        913 GAA---------AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
              :::         :::::::::      ::::::   :::::::::|||   :::
Db        595 LysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGluLysArgArgGln 614

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA--- 1020
              :::   :::         ||||||   |||   :::   :::                |||
Db        615 GluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnValGlnArg 634

Qy       1021 ------------------------------------------------------ATAAGT 1026
                                                                    :::
Db        635 GlnGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysGluLeuGln 654

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              :::   |||:::   :::      |||      :::   |||:::   ||||||   :::
Db        655 ArgGlnGluGluArgLeuGlnGluGlu------GluGlnLeuLeuArgGluGluArgGlu 672

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                    :::      :::      ::::::          |||:::||||||
Db        673 LysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 692

Qy       1147 AACATA---------------------------------------GGAATTTATAAAGAA 1167
                 :::                                          |||   |||
Db        693 GlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGlu 712

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              |||:::|||      :::   |||   |||            |||   :::::::::
Db        713 LysGluGluLeuGlnArgGlnGluGluArgLeuGlnArgGluLysGluGlnLeuGlnArg 732

Qy       1228 TCATCAGAAGAA----------------------------TTATCTGAAGAAAAAATA 1257
                 |||                              |||   ||||||
Db        733 GluAspArgGluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGluLeuGln 752

Qy       1258 AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
              ::::::      :::   ::::::   |||            :::      :::|||
Db        753 GlnGluGluAspArgLeuGlnArgGluLysGlnLeuLeuArgGluAspArgGluLysArg 772
```

```
Qy       1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
Db        773 GlnTyrLeuGluLysValGluLeuGlnArgGluGluGluGlnLeuGlnArgGluLysArg 792

Qy       1378 AAA 1380
Db        793 Arg 793
```

RESULT 4
A45973
trichohyalin - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: A45973
R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A;Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A;Reference number: A45973; MUID:93280194; PMID:7685034
A;Accession: A45973
A;Molecule type: DNA
A;Residues: 1-1898 <LEE>
A;Cross-references: UNIPROT:Q07283; UNIPARC:UPI000013738B; GB:L09190; NID:g292835; PIDN:
A;Note: authors translated the codon AGG for residue 1714 as Pro
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Gene: GDB:THH
A;Cross-references: GDB:136223; OMIM:190370
A;Map position: 1q21-1q21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.:	5.43e-13	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	1	Gaps:	24

US-09-837-344-41 (1-1482) x A45973 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG----------------CAAGAACAACAAAGCGATTTA--- 378
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA----------CAAAGC 474
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 ----------------------------------GATTTAGAACAAGAG--------- 540
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---------------------- 615
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATATCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
```

```
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 5
T15598
hypothetical protein C25A11.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15598
R;Favello, T.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid C25A11.
A;Reference number: Z18375
A;Accession: T15598
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1017 <FAV>
A;Cross-references: UNIPARC:UPI000017B7BA; EMBL:U39650; NID:g1049376; PID:g1049380; PIDN
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C25A11.4a
A;Introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2; 735/1; 792/3; 833/3; 873/2

Alignment Scores:
Pred. No.:              6.73e-12        Length:         1017
Score:                  356.00          Matches:        120
Percent Similarity:     47.80%          Conservative:   141
Best Local Similarity:  21.98%          Mismatches:     203
Query Match:            14.32%          Indels:         82
DB:                     2               Gaps:           18

US-09-837-344-41 (1-1482) x T15598 (1-1017)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        226 GluGluGluAlaGluLeuAlaArg---ArgArgAlaLeuGluLysGluArgIleAspArg 244

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
Db        245 GluLysAlaGluGluGluArgLysThrMetGluArgLeuGluArgGluArgAlaArgLeu 264

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------CTA 174
Db        265 GluArgGluArgLeuGluGluGluArgArgGlnLysGluLysAlaGluThrGluArgIle 284

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAA 231
Db        285 GluArgGluArgArgGluHisGluArgIleGluIleGluArgIleGluArgIleLysArg 304

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
Db        305 GluArgIleGluArgGluArgArgGluArgGluGluLysLysAlaGluGluAspArgLeu 324

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
Db        325 LeuArgGluArgLeuGluLeuGluArg------IleGluArgGluArgArgGluLeuGlu 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACGTGCTAAAGAA 402
Db        343 AlaArgGluArgGlnGluLeuGluLeuGlnArgArgGluAlaGluAspArgGluArgGln 362

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACTT--------- 444
Db        363 ArgLeuGluAspGluAlaArgGluMetArgArgArgGluGluGluArgArgGluAlaGlu 382

Qy        445 ----------------------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
Db        383 LeuValAlaAspValHisArgGlnAlaGluGluArgGluArgLeuArgLysArgGlnGlu 402

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
Db        403 ArgGluGluAlaGluArgLeuGluArgIleArgLeuGluGlnGlnLysIleAspMetGlu 422

Qy        535 CAA---------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
Db        423 ArgIleAspAlaGluArgArgGluArgGluArgLysGluGluGluArgArgGluPheGlu 442

Qy        586 ----------------------CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
Db        443 LeuIleGluAlaAlaArgArgLysLysGluAlaArgAspArgAspArgLeuAspGluMet 462

Qy        622 CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
Db        463 GluArgGlu---ArgValArgGluGluGluGluArgArgGluLysGluArgArgGluGln 481

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
Db        482 GluArgArgIleAlaAlaGluLysGluArgLysArgArgGlnGluGluGluGluGluIle 501

Qy        742 CCATCAGAAAATGAA--------------CGTGGATATTATATACCACATCAATCTTCT 786
Db        502 AlaArgLeuAsnGluLeuGlnArgAlaAlaAlaAlaArgGlnAlaGlnArgAsnAlaGlu 521

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAA 846
Db        522 LeuAspArgGlnArgGlnArgAspGluLeuAspArgLysAlaGlnGluLeuSerGluArg 541

Qy        847 ACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACAT 906
Db        542 GluMetArgGlu-----------------LysGluArgArgAspArgGluArgAlaAsn 555

Qy        907 CTTGAA--------------------------GAAAAGAAAGATGGTTCAATAAAACCA 939
Db        556 GluGluAlaGlnLeuAlaAspLeuLeuGluArgGluArgHisAsnGlnLeuIleArgGlu 575

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
Db        576 AsnGluArgArgGluAlaValGluArgAlaAsnAsnArgArgLeuGluAspArgArgSer 595

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
Db        596 ArgAspLysLeuAspHisIleValArgGluArgSerGluLysGluGlnPheGluLeuGlu 615

Qy       1060 GAT---TCATTAATAGATGAAGAAGAA----------------------------GAT 1086
Db        616 LysArgArgLeuLeuAlaGluLysGluAlaMetAsnArgLysLysAsnHisLeuLeuSer 635

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
Db        636 SerGluThrLeuAlaLysLeuThrGlnProMetTyrTyrThrThrArgGluProGluVal 655

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT 1206
Db        656 ThrThrLysValGluArgGlnValIleGluArgIleAspArgAsnValTrpValGluAsp 675

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
Db        676 ValProTyrAlaProSerGlnSerAlaMetGlyTyrLeuAspAsnAsp------------ 691

Qy       1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
Db        692 ---------GluAsnAsnArgAspArgLeuTyrAsnProAsnAspLeuAsnArgAsnGly 708

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
Db        709 SerSerArgSerArgTyrGlnArgAlaLys--------------AsnGluLysAlaArg 723

Qy       1387 AAATCATTGTTTCATATATTT---GACGGAGACAATGAAATTTTACAGATCGTGGATGAG 1443
Db        724 ArgAspPheTyrHisSerSerGlnAspSerAlaAspProValThrGluArgPheArgLys 743

Qy       1444 TTATCTGAAGATATAACT 1461
```

```
Db        744 SerThrAspAspLeuThr 749

RESULT 6
T15597
hypothetical protein C25A11.4b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15597
R;Favello, T.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid C25A11.
A;Reference number: Z18375
A;Accession: T15597
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-737 <FAV>
A;Cross-references: UNIPARC:UPI000017B7BB; EMBL:U39650; NID:g1049376; PID:g1049381; PIDN
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C25A11.4b
A;Introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2

Alignment Scores:
Pred. No.:              9.6e-12        Length:          737
Score:                  353.50         Matches:         117
Percent Similarity:     47.81%         Conservative:    134
Best Local Similarity:  22.29%         Mismatches:      193
Query Match:            14.22%         Indels:          81
DB:                     2              Gaps:            17

US-09-837-344-41 (1-1482) x T15597 (1-737)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||:::::::::::::|||   :::   |||||||||   ||||||   :::      :::
Db        226 GluGluGluAlaGluLeuAlaArg---ArgArgAlaLeuGluLysGluArgIleAspArg 244

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                  |||::::::|||         |||:::|||:::   :::::::::   |||
Db        245 GluLysAlaGluGluGluArgLysThrMetGluArgLeuGluArgGluArgAlaArgLeu 264

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------CTA 174
              |||:::|||||||||   :::|||:::   |||:::::::::::::::         :::
Db        265 GluArgGluArgLeuGluGluGluArgArgGlnLysGluLysAlaGluThrGluArgIle 284

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAA 231
              |||:::|||||||||      |||:::::::::   |||:::         :::::::::
Db        285 GluArgGluArgArgGluHisGluArgIleGluIleGluArgIleGluArgIleLysArg 304

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
              ||||||      :::|||:::   :::   ::::::         |||:::|||||||||
Db        305 GluArgIleGluArgGluArgArgGluArgGluGluLysLysAlaGluGluAspArgLeu 324

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::|||:::|||:::   ::::::      :::|||:::|||||||||      |||
Db        325 LeuArgGluArgLeuGluLeuGluArg------IleGluArgGluArgArgGluLeuGlu 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACGTGCTAAAGAA 402
                  :::   |||:::   :::|||          |||   |||   |||   ::::::
Db        343 AlaArgGluArgGlnGluLeuGluLeuGlnArgArgGluAlaGluAspArgGluArgGln 362

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACTT--------- 444
              :::|||::::::::::      ::::::      |||:::||||||
Db        363 ArgLeuGluAspGluAlaArgGluMetArgArgArgGluGluGluArgArgGluAlaGlu 382

Qy        445 ----------------------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
                                         :::|||:::|||::::::::::|||
Db        383 LeuValAlaAspValHisArgGlnAlaGluGluArgGluArgLeuArgLysArgGlnGlu 402

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                  |||   ||||||      :::   :::|||::::::|||:::   |||:::|||
Db        403 ArgGluGluAlaGluArgLeuGluArgIleArgLeuGluGlnGlnLysIleAspMetGlu 422

Qy        535 CAA---------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                  :::         |||||||||   :::|||:::   :::|||::::::   :::   |||
Db        423 ArgIleAspAlaGluArgArgGluArgGluArgLysGluGluGluArgArgGluPheGlu 442

Qy        586 ----------------------CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
                                     :::|||   |||   :::::::::|||   |||
Db        443 LeuIleGluAlaAlaArgArgLysLysGluAlaArgAspArgAspArgLeuAspGluMet 462

Qy        622 CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::|||:::         |||:::   :::   ::::::      ||||||:::::::::
Db        463 GluArgGlu---ArgValArgGluGluGluGluArgArgGluLysGluArgArgGluGln 481

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
                    |||   ||||||              |||   :::             |||:::
Db        482 GluArgArgIleAlaAlaGluLysGluArgLysArgArgGlnGluGluGluGluGluIle 501

Qy        742 CCATCAGAAAATGAA---------------CGTGGATATTATATACCACATCAATCTTCT 786
                    ||||||                                          :::
Db        502 AlaArgLeuAsnGluLeuGlnArgAlaAlaAlaAlaArgGlnAlaGlnArgAsnAlaGlu 521

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAA 846
              |||   :::       :::   :::       |||   |||          :::   |||:::
Db        522 LeuAspArgGlnArgGlnArgAspGluLeuAspArgLysAlaGlnGluLeuSerGluArg 541

Qy        847 ACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACAT 906
                    ||||||                  :::   |||||||||       :::   :::
Db        542 GluMetArgGlu------------------LysGluArgArgAspArgGluArgAlaAsn 555

Qy        907 CTTGAA--------------------------GAAAAGAAAGATGGTTCAATAAAACCA 939
                 |||                         |||:::   :::      |||:::
Db        556 GluGluAlaGlnLeuAlaAspLeuLeuGluArgGluArgHisAsnGlnLeuIleArgGlu 575

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                  ::::::   :::                |||       ||||||
Db        576 AsnGluArgArgGluAlaValGluArgAlaAsnAsnArgArgLeuGluAspArgArgSer 595

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 |||       |||      :::   :::      ::::::          |||   :::
Db        596 ArgAspLysLeuAspHisIleValArgGluArgSerGluLysGluGlnPheGluLeuGlu 615

Qy       1060 GAT---TCATTAATAGATGAAGAAGAA----------------------------GAT 1086
                    |||:::   |||:::|||
Db        616 LysArgArgLeuLeuAlaGluLysGluAlaMetAsnArgLysLysAsnHisLeuLeuSer 635

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                 |||   |||   :::      :::   |||          :::   |||
Db        636 SerGluThrLeuAlaLysLeuThrGlnProMetTyrTyrThrThrArgGluProGluVal 655

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT 1206
                    :::   :::::::::   :::   |||::::::|||      ::::::|||
Db        656 ThrThrLysValGluArgGlnValIleGluArgIleAspArgAsnValTrpValGluAsp 675

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
                 :::                :::|||:::            :::
Db        676 ValProTyrAlaProSerGlnSerAlaMetGlyTyrLeuAspAsnAsp------------ 691

Qy       1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
                    |||      :::|||      :::   |||||||||   :::
Db        692 ---------GluAsnAsnArgAspArgLeuTyrAsnProAsnAspLeuAsnArgAsnGly 708

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
                       :::|||:::      |||                  ||||||
Db        709 SerSerArgSerArgTyrGlnArgAlaLys--------------AsnGluLysAlaArg 723

Qy       1387 AAATCATTGTTTCAT 1401
                 :::      :::|||
Db        724 ArgAspPheTyrHis 728

RESULT 7
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: A71623
```

R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1192 <GAR>
A;Cross-references: UNIPROT:O96127; UNIPARC:UPI0000076661; GB:AE001373; GB:AE001362; NID
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0115w

Alignment Scores:

Pred. No.:	1.11e-10	Length:	1192
Score:	332.50	Matches:	112
Percent Similarity:	46.85%	Conservative:	126
Best Local Similarity:	22.05%	Mismatches:	193
Query Match:	13.37%	Indels:	77
DB:	2	Gaps:	16

US-09-837-344-41 (1-1482) x A71623 (1-1192)

```
Qy         10 CAAAGCGATCTAGAACAAGAGAGA--------------------------------CGT  36
              ||||||      :::|||||||||                                  :::
Db        296 GlnSerLysTyrLysGlnGluArgIleGluIleLeuAspAspAsnGlyLysGluLeuLys  315

Qy         37 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA---------GATAGACTT  87
              :::   :::   ::::::|||::::::      :::|||:::               :::
Db        316 SerHisLysAsnIleLysGluGluLysGlyGlyIleGluLysThrAspThrThrAsnIle  335

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA  147
              |||      |||:::::::::::::::   :::   :::   |||:::      ::::::
Db        336 AlaAspIleLysIleLysLysGluGluArgGluThrLysAspGluLys------GluLys  353

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA  207
              :::|||:::         |||::::::      :::      :::|||
Db        354 AsnIleGlnGlnLeuValLysAspValGlnLeuIleLysValGlyGluGluThrLysAsp  373

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA  267
              :::::::::   |||   |||                 ::::::         |||:::
Db        374 AspGluLysGluAspLysGluGlyThrAspAspGluGluAspThrAspAspGluGluAsp  393

Qy        268 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA  327
              :::|||   |||:::|||       :::|||      ::::::::::::
Db        394 ThrAspAspGluGluAspThrAspAspGluGluAspThrSerAspGluGluThrThrGly  413

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG  387
              :::||||||   :::   :::   ::::::   |||::::::::::::
Db        414 AspGlnGluAsnLysGluGluThrGluValAspGluLysLysThrGlu------------  429

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT  447
              :::|||:::|||:::|||:::|||   :::   :::   |||:::::::::
Db        430 ---LysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGluGlu  448

Qy        448 AAAGAAAAG---TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA  504
              ||||||      :::|||   :::   |||   |||:::      :::   ||||||
Db        449 SerGluLysAspLysGluGluSerGluLysAspLysGluGluSerGluLysAspLysGlu  468

Qy        505 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA  564
              |||   :::|||   :::           |||   |||:::      :::          :::
Db        469 LysThrGluGluAspGluGluLysThrGluAspGluLysGlyThrGluValTyrLysLys  488

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAA  624
              |||          |||   ::::::|||:::           :::      :::   :::
Db        489 GluThrAspValAspGluLysLysGluLysGlyGluTyrGlyGluGlyThrAspAspGlu  508

Qy        625 AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACAT  684
              |||   |||:::   :::   |||   :::      :::|||   ||||||
Db        509 GluAspLysGluLysGluGluAspAspGluGluThrLysValGluGluLysLysThrGlu  528

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCA  744
              |||
Db        529 LysAsp------------------------------------------------------  530

Qy        745 TCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA  804
                  |||   |||                 :::        :::|||
Db        531 ---------GluGluGly-----------------ThrAspTyrGluGluAspThrAsp  541

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGA---GAATCT  861
              :::   :::|||   :::               :::||||||   |||   |||
Db        542 AspSerAspLysAspGluGluThrLysValGluGluLysLysThrGluArgAspGluGlu  561

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA  921
              |||         |||       ::::::   |||       |||:::   :::
Db        562 GluThrGluAspGluLysGluThrGluValGluLysLysLysThrGluLysAspGlu  581

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA  981
              :::|||:::         |||      :::|||         |||::::::
Db        582 GluGlyThrAspTyrGluGluAspThrAspAspSerAspLysAspValGluThrGluVal  601

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAG 1035
              ||||||              |||      |||   :::   ::::::         :::
Db        602 GluGluThr------------AspAlaGluAspLysGluGluAsnGluGluGlyThrAsp  617

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT---GATGAA 1092
              ||||||                :::      |||   |||:::||||||||   :::|||
Db        618 AspGluGluAspLysValGluGluThrAspLeuAspAspGlnGluGluAspGlyGluGlu  637

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
              |||   ::::::   |||         :::   |||:::   :::|||::::::
Db        638 AspLysGluAspAspLysGluLysAspLysGluAspAspLysGluAspAspLysGlu--- 656

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA--------------------AAT 1191
              |||:::   ||||||      ||||||                           :::
Db        657 ---------LysAspLysGluAspAspLysGluLysTyrLysGluAspAspLysGluAsp  673

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
              :::      ||||||   ::::::      |||   :::   |||:::      |||:::
Db        674 AspLysGluAspAspLysGluLysAspLysGluAspAsnLysGluLysAspLysGluAsp  693

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              :::|||   |||:::   :::|||   ||||||         |||   :::|||
Db        694 AsnLysGluLysAspLysGluAspAspLysGluLysAsp---------LysGluAspAsp  710

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
              |||         |||      |||      :::   ||||||:::   |||||||||
Db        711 LysGluLysAspLysGluAspAsnLys------GluLysAspLysGluAspAsnLysGlu  728

Qy       1372 AAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
              |||:::|||                |||            |||      :::   :::
Db        729 LysAspLysGluAspAspLysGluLysHis-----------AspLysHisValArgArg  744

Qy       1432 ATCGTGGATGAGTTATCTGAAGAT 1455
              |||      ::::::   :::|||
Db        745 IleLysLysLysMetLysAspAsp  752
```

RESULT 8
S03166
myosin heavy chain, gizzard smooth muscle [similarity] - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S03166; A27066; A26045; A36604; A43298
R;Yanagisawa, M.; Hamada, Y.; Katsuragawa, Y.; Imamura, M.; Mikawa, T.; Masaki, T. J. Mol. Biol. 198, 143-157, 1987
A;Title: Complete primary structure of vertebrate smooth muscle myosin heavy chain deduc
A;Reference number: S03166; MUID:88118918; PMID:2892941
A;Accession: S03166
A;Molecule type: mRNA
A;Residues: 1-1979 <YAN>
A;Cross-references: UNIPROT:P10587; UNIPARC:UPI0000171370; EMBL:X06546; NID:g63633; PIDN
A;Note: part of this sequence was confirmed by protein sequencing
R;Maita, T.; Onishi, H.; Yajima, E.; Matsuda, G. J. Biochem. 102, 133-145, 1987

```
A;Title: Amino acid sequence of the amino-terminal 24 kDa fragment of the heavy chain of
A;Reference number: A27066; MUID:88032919; PMID:3312184
A;Accession: A27066
A;Molecule type: protein
A;Residues: 2,'Z',4-204 <MAI>
A;Cross-references: UNIPARC:UPI0000173E2B
R;Onishi, H.; Maita, T.; Miyanishi, T.; Watanabe, S.; Matsuda, G.
J. Biochem. 100, 1433-1447, 1986
A;Title: Amino acid sequence of the 203-residue fragment of the heavy chain of chicken g
A;Reference number: A26045; MUID:87194651; PMID:3571180
A;Accession: A26045
A;Molecule type: protein
A;Residues: 653-855 <ONI>
A;Cross-references: UNIPARC:UPI0000173E2C
R;Onishi, H.; Maita, T.; Matsuda, G.; Fujiwara, K.
J. Biol. Chem. 265, 19362-19368, 1990
A;Title: Lys-65 and Glu-168 are the residues for carbodiimide-catalyzed cross-linking be
A;Reference number: A36604; MUID:91035476; PMID:1977747
A;Accession: A36604
A;Status: preliminary
A;Molecule type: protein
A;Residues: 54-67;146-183 <ON2>
A;Cross-references: UNIPARC:UPI0000173E2D; UNIPARC:UPI0000173E2E
R;Cole, D.G.; Yount, R.G.
Biochemistry 31, 6186-6192, 1992
A;Title: Stability and photochemical properties of vanadate-trapped nucleotide complexes
A;Reference number: A43298; MUID:92329440; PMID:1385724
A;Accession: A43298
A;Status: preliminary
A;Molecule type: protein
A;Residues: 169-183 <COL>
A;Cross-references: UNIPARC:UPI0000173E2F
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; blocked amino end; coiled coil; hydrolase; methylated am
F;87-777/Domain: myosin motor domain homology <MMOT>
F;177-184/Region: nucleotide-binding motif A (P-loop)
F;565-578/Region: actin binding #status predicted
F;639-653/Region: actin binding #status predicted
F;850-1940/Domain: coiled coil <COI>
F;850-1290/Region: S2
F;1291-1979/Region: light meromyosin
F;1941-1979/Domain: carboxyl-terminal <CBT>
F;2/Modified site: blocked amino end (Ser) (in mature form) #status experimental
F;128/Modified site: N6,N6,N6-trimethyllysine (Lys) #status experimental
F;183/Binding site: ATP (Lys) #status predicted
F;707,717/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                2.11e-10      Length:          1979
Score:                    326.50        Matches:         154
Percent Similarity:       42.17%        Conservative:    134
Best Local Similarity:    22.55%        Mismatches:      186
Query Match:              13.13%        Indels:          209
DB:                       1             Gaps:            35

US-09-837-344-41 (1-1482) x S03166 (1-1979)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::        :::|||   :::      |||:::|||:::|||:::|||
Db        938 GluGluGluGluGluArgSerGlnGlnLeuGlnAlaGluLysLysLysMetGlnGlnGln 957

Qy         61 CAAAGCGAT-----------TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA--- 105
                 |||          ||||||::::::      |||::::::|||||||||
Db        958 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 977

Qy        106 ----------------------------GAGCAGCAAAGCGATTTA------------ 123
                                          :::   :::   |||:::
Db        978 GluLysValThrAlaAspGlyLysIleLysLysMetGluAspAspIleLeuIleMetGlu 997

Qy        124 ---GAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                  :::      :::|||   ||||||:::      |||:::|||:::   |||||||||
Db        998 AspGlnAsnAsnLysLeuThrLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1017

Qy        175 --------------GAACAAGAGAGACGTGCTAAAGAA---------------------- 198
                             |||:::|||   :::||||||
Db       1018 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1037

Qy        199 ----------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAA 228
                                         :::|||::::::::::::          :::
Db       1038 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1057

Qy        229 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT--- 285
               :::      :::   |||   ||||||:::   :::   |||||||||   :::
Db       1058 GluLeuGluLysIleLysArgLysLeuGluGlyGluSerSerAspLeuHisGluGlnIle 1077

Qy        286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
                                               :::|||||||||      |||:::
Db       1078 AlaGluLeuGlnAlaGlnIleAlaGluLeuLysAlaGlnLeuAlaLysLysGluGluGlu 1097

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG-------------------------- 336
              ||||||         :::   ||||||   |||
Db       1098 LeuGlnAlaAlaLeuAlaArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLys 1117

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
                    |||      :::      :::   :::   |||   |||||||||   |||:::
Db       1118 LysIleArgGluLeuGluSerHisIleSerAspLeuGlnGluAspLeuGluSerGluLys 1137

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                ||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1138 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAlaLeu 1157

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::            :::      :::||||||   |||||||||
Db       1158 LysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLys 1177

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||:::|||:::           |||
Db       1178 ArgGluGlnGluValThrValLeuLysArgAlaLeuGluGluGluThrArgThrHisGlu 1197

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1198 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1217

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1218 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLys 1237

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                 ||||||                              |||::::::::::|||
Db       1238 AspAsnAlaAspLeuAlaAsnGluIleArgSerLeuSerGlnAlaLysGlnAspValGlu 1257

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA--------- 720
                 ||||||:::      ::::::      :::||||||      :::
Db       1258 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysTyrSerAspGly 1277

Qy        721 --------------------------GAAATACCAGCTATAGAACTTCCATCAGAAAAT 753
                                        ::::::      :::::::::      ||||||
Db       1278 GluArgValArgThrGluLeuAsnGluLysValHisLysLeuGlnIleGluValGluAsn 1297

Qy        754 GAACGTGGATATTATATACCACATCAATCT---------TCTTTACCTCAGGAC------ 798
                                       :::|||           |||   :::|||
Db       1298 ValThrSerLeuLeuAsnGluAlaGluSerLysAsnIleLysLeuThrLysAspValAla 1317

Qy        799 AACAGAGGGAATAGT---AGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGA 855
                ||:::      :::|||::::::|||:::         |||          :::
Db       1318 ThrLeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLys 1337

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
               ::::::||||||   :::          |||      |||      |||:::|||
Db       1338 LeuAsnValThrThrLysLeuArgGlnLeuGluAsp---AspLysAsnSerLeuGlnGlu 1356

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
               :::   |||      ::::::   :::|||                 :::::::::
Db       1357 GlnLeuAspGluGluValGluAlaLysGln-----------------AsnLeuGluArg 1370

Qy        976 CAT------ACATTAGAGACAGTAAATATTTCTGAT------------GTTAATGATTTT 1017
```

```
Db       1371 HisIleSerThrLeu---ThrIleGlnLeuSerAspSerLysLysLysLeuGlnGluPhe 1389

Qy       1018 ------------------------------------------------------------CAA 1020

Db       1390 ThrAlaThrValGluThrMetGluGluGlyLysLysLysLeuGlnArgGluIleGluSer 1409

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC---------------GATTCA 1065

Db       1410 LeuThrGlnGlnPheGluGluLysAlaAlaSerTyrAspLysLeuGluLysThrLysAsn 1429

Qy       1066 TTAATAGATGAAGAAGAAGATGAT------GAAGACTTAGACGAATTTAAGCCTATTGTG 1119

Db       1430 ArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGlnArgGlnLeuVal 1449

Qy       1120 ----------------------CAATATGACAATTTCCAAGATGAAGAAAAC------ 1149

Db       1450 SerAsnLeuGluLysLysGlnLysLysPheAspGlnMetLeuAlaGluGluLysAsnIle 1469

Qy       1149 ------------------------------------------------------------ 1149

Db       1470 SerSerLysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThr 1489

Qy       1150 ------ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203

Db       1490 LysAlaLeuSerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeuGlu 1509

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263

Db       1510 ArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLysAspAspVal 1529

Qy       1264 GGAAAG---------AAATATGAAAAAACAAAGGAT-----------------AATAAT 1296

Db       1530 GlyLysAsnValHisGluLeuGluLysSerLysArgThrLeuGluGlnGlnValGluGlu 1549

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356

Db       1550 MetLysThrGlnLeuGluGluLeuGluAspGlu---LeuGlnAlaAlaGluAspAlaLys 1568

Qy       1357 ---------CAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407

Db       1569 LeuArgLeuGluValAsnMetGlnAla---------MetLysSerGlnPhe--------- 1582

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467

Db       1583 ---------GluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArgArgGln 1599

Qy       1468 TTTATGAAA 1476

Db       1600 LeuLeuLys 1602

RESULT 9
S18199
myosin heavy chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 02-Feb-2001
C;Accession: S18199
R;Stewart, A.F.R.; Camoretti-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.
J. Mol. Evol. 33, 357-366, 1991
A;Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy ch
A;Reference number: S18199; MUID:92130260; PMID:1774788
A;Accession: S18199
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1039 <STE>
A;Cross-references: UNIPARC:UPI0000171256; EMBL:X59552; NID:g62995; PIDN:CAA42130.1; PID
A;Note: in the authors' translation 45-Lys is shown after residue 40, and, consequently,
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: ATP; nucleotide binding; P-loop
F;402-409/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.:              3.37e-10        Length:         1039
```

```
Score:                  323.50          Matches:        143
Percent Similarity:     40.99%          Conservative:   114
Best Local Similarity:  22.81%          Mismatches:     187
Query Match:            13.01%          Indels:         183
DB:                     2               Gaps:           27

US-09-837-344-41 (1-1482) x S18199 (1-1039)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---------- 45

Db         32 GluGluMetAsnSerGluLeuThrSerLysLysArgLysLeuGluAspGluPheAlaLeu 51

Qy         46 AAGTTGCAAGAACAACAAAGCGATTTAGAA------CAAGATAGACTTGCTAAAGAAAAG 99

Db         52 GluLeuLysLysAspIleAspAspLeuGluIleThrLeuAlaLysValGluLysGluLys 71

Qy        100 --------------------------TTACAAGAGCAGCAAAGCGATTTAGAACAA--- 129

Db         72 GlnAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaThrLeuAspGluAsn 91

Qy        130 ---GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAA------AGC 168

Db         92 IleSerLysLeuThrLysGluLysLysSerLeuGlnGluAlaHisGlnGlnValLeuAsp 111

Qy        169 GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAGTTGCAA 207

Db        112 AspLeuGlnAlaGluGluAspLysValAsnArgLeuSerLysAlaLysValLysLeuGlu 131

Qy        208 GAACAACAAAGCGAT------------TTAGAACAAGAGAGA----------------- 237

Db        132 GlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysValArgMetAspLeu 151

Qy        238 ---CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAAAGCGAT 273

Db        152 GluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuThrGlnGluSerValMetAsp 171

Qy        274 TTAGAACAAGATAGACTT---GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330

Db        172 LeuGluAsnAspLysLeuGlnMetGluGluLysLeuLysLysLysGluPheGluMetSer 191

Qy        331 CAA-------------------------------------------GAGAGACGTGCT 345

Db        192 GlnLeuAsnSerLysIleGluAspGluGlnAlaIleValMetGlnLeuGlnLysLysIle 211

Qy        346 AAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 396

Db        212 LysGluLeuGlnAlaArgIleGluGluLeuGluGluGluLeuGluAlaGluArgAlaAla 231

Qy        397 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 456

Db        232 ArgAlaLysValGluLysGlnArgSerAspLeuAlaArgGlu---------LeuGluGlu 248

Qy        457 TTGCAAGAACAA----------------------CAAAGCGATTTAGAACAAGAGAGA 492

Db        249 LeuSerGluArgLeuGluGlyAlaGlyGlyAlaThrAlaAlaGlnLeuGluMetAsnLys 268

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA---------------- 537

Db        269 LysArgGluAlaGluPheLeuLysLeuAlaArgAspLeuGluGluAlaThrLeuHisTyr 288

Qy        538 --------------------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 576

Db        289 GluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspArgArgGlyGluMetGlyGlu 308

Qy        577 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAA 636

Db        309 GlnLeuAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLys 328

Qy        637 CAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696

Db        329 MetGluValAspAspLeuThrAlaAsnMetGluGlnThrValLys---GlyLysAlaAsn 347

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAA------------------------------ 723
```

```
Db        348 AlaGluLysLeuTrpGlyThrTyrGluAspHisLeuAsnGluThrLysThrLysLeuAsp 367

Qy        724 ----------------------ATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Db        368 GluMetThrArgLeuMetAsnAspLeuThrThrGlnLysThrLysLeuGlnSerGluAsn 387

Qy        760 GGATATTATATA------------CCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807

Db        388 GlyGluPheValArgGlnLeuGluGluLysGluSerLeuIleSerGlnLeuSerArgGly 407

Qy        808 AATAGTAGAGATTCCAAGGAAATA---------------------------------- 831

Db        408 LysThrSerPheThrGlnGlnIleGluGluLeuArgArgGlnLeuGluGluGluThrLys 427

Qy        832 ----------------------------------------------TCTATAATAGAA 843

Db        428 SerLysAsnAlaLeuAlaHisAlaLeuGlnAlaAlaArgHisAspCysAspLeuLeuArg 447

Qy        844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903

Db        448 GluGlnTyrGluGluGluGlnGluAlaLysAlaGluLeuGlnArgAlaLeuSerLysGly 467

Qy        904 CATCTTGAA--------------------GAAAAGAAAGATGGTTCAATAAAACCAGAA 942

Db        468 AsnAlaGluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGlnArgThrGlu 487

Qy        943 CAAAAAGAAGATAAATCTGCTGACATA---------CAAAATCATACATTAGAGACAGTA 993

Db        488 GluLeuGluAspAlaLysLysLysLeuLeuAlaArgLeuGlnGluAlaGluGluAlaIle 507

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053

Db        508 GluAlaAlaAsnAlaLysCysSerSerLeuGluLysAlaLysHisArgLeuGlnAsnGlu 527

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113

Db        528 GlnGluAspMetMetIleAspLeuGluLysAlaAsnSerAlaAlaAlaSerLeu------ 545

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173

Db        546 ----------------------AspLysLysGlnArgGlyPheAspLysIleIleAsn 557

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233

Db        558 AspTrpLysGlnLysTyrGluGluSerGlnAlaGluLeuGluAlaSerGlnLysGluAla 577

Qy       1234 GAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293

Db        578 ArgSerLeuSerThrGluLeuPheLysLeuLysAsnAlaTyrGluGluThrLeuAspHis 597

Qy       1294 ------AATTTTAAACCAAATGATAAAAGTTTGTATGATGAG----------------- 1329

Db        598 LeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGluGluIleSerAspLeuThrAsn 617

Qy       1330 ----------------------------CATATTAAAAAATATAAAAATGATAAGCAG 1359

Db        618 GlnIleSerGluGlyAsnLysAsnLeuHisGluIleGluLysValLys------LysGln 635

Qy       1360 GTTAATAAGGAAAAGGAAAAA 1380

Db        636 ValGluGlnGluLysSerGlu 642
```

RESULT 10
A33430
h-caldesmon - chicken
N;Alternate names: caldesmon, smooth muscle; calmodulin- and actin-binding protein
C;Species: Gallus gallus (chicken)
C;Date: 27-Feb-1990 #sequence_revision 27-Jun-1994 #text_change 09-Jul-2004
C;Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022
R;Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 164, 503-511, 1989
A;Title: Primary structure and functional expression of h-caldesmon complementary DNA.
A;Reference number: A33430; MUID:90026426; PMID:2803315
A;Accession: A33430
A;Molecule type: mRNA
A;Residues: 1-771 <HAY>
A;Cross-references: UNIPROT:P12957; UNIPARC:UPI0000126E45; GB:M28417; NID:g211895; PIDN:
A;Experimental source: gizzard
A;Note: part of this sequence was confirmed by protein sequencing
R;Bryan, J.; Imai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.
J. Biol. Chem. 264, 13873-13879, 1989
A;Title: Cloning and expression of a smooth muscle caldesmon.
A;Reference number: A32642; MUID:89340480; PMID:2760048
A;Accession: A32642
A;Molecule type: mRNA
A;Residues: 1-318,334-771 <BRY>
A;Cross-references: UNIPARC:UPI00001713AB; GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:
A;Note: the authors translated the codon GAA for residue 743 as Leu
A;Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)
R;Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 161, 38-45, 1989
A;Title: 35kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo
A;Reference number: A32445; MUID:89273666; PMID:2730665
A;Accession: A32445
A;Molecule type: mRNA
A;Residues: 466-771 <HA2>
A;Cross-references: UNIPARC:UPI000015624D; GB:M26684; NID:g211897; PIDN:AAA48811.1; PID:
R;Mak, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.
J. Biol. Chem. 266, 19971-19975, 1991
A;Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphoryla
A;Reference number: A41064; MUID:92041815; PMID:1939059
A;Accession: A41064
A;Molecule type: protein
A;Residues: 597-600;678-696;711-721 <MAK>
A;Cross-references: UNIPARC:UPI0000173E63; UNIPARC:UPI0000173E64; UNIPARC:UPI0000173E65
R;Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.
Biochem. Biophys. Res. Commun. 162, 746-752, 1989
A;Title: Caldesmon has two calmodulin-binding domains.
A;Reference number: A60461; MUID:89334885; PMID:2757638
A;Accession: A60461
A;Molecule type: protein
A;Residues: 2-17,'X',19-38;466-485 <WAN>
A;Cross-references: UNIPARC:UPI0000173E66; UNIPARC:UPI0000173E67
R;Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.
Biochem. Biophys. Res. Commun. 197, 145-153, 1993
A;Title: Common structural and expressional properties of vertebrate caldesmon genes.
A;Reference number: PC2003; MUID:94071934; PMID:8250919
A;Accession: PC2003
A;Molecule type: DNA
A;Residues: 74-419 <HAR>
A;Cross-references: UNIPARC:UPI0000173E68
R;Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.
J. Biochem. 106, 778-783, 1989
A;Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon w
A;Reference number: PX0022; MUID:90130380; PMID:2613684
A;Accession: PX0022
A;Molecule type: protein
A;Residues: 462-477,'D',479-563;674-762,'A',763-771 <TAK>
A;Cross-references: UNIPARC:UPI0000173E69; UNIPARC:UPI0000173E6A
C;Comment: This protein plays a vital role in the regulation of smooth muscle and nonmus
C;Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C;Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C;Superfamily: caldesmon
C;Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; muscl
F;1-771/Product: h-caldesmon #status predicted <HMAT>
F;1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>
F;266-390/Region: 13-residue repeats
F;511-582/Region: tropomyosin binding
F;622-636/Region: tropomyosin binding
F;597,682,717/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status experime
F;688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimental

Alignment Scores:
Pred. No.: 4.8e-10 Length: 771
Score: 321.00 Matches: 122

```
Percent Similarity:     47.07%          Conservative:  111
Best Local Similarity:  24.65%          Mismatches:    172
Query Match:            12.91%          Indels:        90
DB:                     1               Gaps:          21

US-09-837-344-41 (1-1482) x A33430 (1-771)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::::::::::   ||| |||:::|||:::     :::         :::|||
Db        179 LysLysGluGluLysAspSerGluGluGluLysProLysGluValProThrGluGluAsn 198

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||  |||              :::|||  ||||||     :::::::::
Db        199 GlnValAsp--------------ValAlaValGluLysSerThrAspLysGluGluVal 213

Qy        121 TTAGAACAAGAGAGACTTGCT--------------------AAAGAAAAGTTGCAAGAA 159
              :::|||  :::  ||||||                     |||:::
Db        214 ValGluThrLysThrLeuAlaValAsnAlaGluAsnAspThrAsnAlaMetLeuGluGly 233

Qy        160 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 219
              :::||||||     :::         :::  ||||||:::  ::::::::::::
Db        234 GluGlnSerIleThrAspAlaAlaAspLysGluLysGluGluAlaGluLysGluArgGlu 253

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
              ||||||  |||  :::  :::  |||  :::|||:::::      |||
Db        254 LysLeuGluAlaGluGluLysGluArgLeuLysAlaGluGluGluLysLysAlaAlaGlu 273

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTA------------CAAGAGCAGCAAAGCGATTTA 327
              :::  :::  |||:::|||:::              :::  :::::::::
Db        274 GluLysGlnLysAlaGluGluGluLysLysAlaAlaGluGluArgGluArgAlaLysAla 293

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              |||:::|||:::||||||  |||      :::  :::::::::      |||:::|||
Db        294 GluGluGluLysArgAlaAlaGlu------GluArgGluArgAlaLysAlaGluGluGlu 311

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
              |||:::|||  |||      :::  :::::::::      |||:::||||||  |||
Db        312 ArgLysAlaAlaGlu------GluArgGluArgAlaLysAlaGluGluGluArgLysAla 329

Qy        448 AAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 498
              |||         |||  :::|||::::::         |||:::      :::|||
Db        330 AlaGluGluArgAlaLysAlaGluGluGluArgLysAlaAlaGluGluArgAlaLysAla 349

Qy        499 AAAGAA------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              :::|||      |||      |||:::::::::      |||:::||||||:::|||
Db        350 GluGluGluArgLysAlaAlaGluGluArgAlaLysAlaGluLysGluArgLysAlaAla 369

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT------------ 600
              |||      :::  :::::::::      |||:::|||:::||||||
Db        370 Glu------GluArgGluArgAlaLysAlaGluGluGluLysArgAlaAlaGluGluLys 387

Qy        601 ------------AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
                         |||||||||:::|||:::::::::   :::|||::::::||||||
Db        388 AlaArgLeuGluAlaGluLysLeuLysGluLysLysLys---MetGluGluLysLysAla 406

Qy        649 GATACGAAAAAA---------AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                  :::|||         ||||||   ||||||:::|||
Db        407 GlnGluGluLysAlaGlnAlaAsnLeuLeuArgLysGlnGlu----------------- 420

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||         ::::::|||              ||||||   :::   :::
Db        421 GluAspLysGluAlaLysValGluAlaLysLysGluSerLeuProGluLysLeuGln--- 439

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
                                          |||      :::      :::|||
Db        440 ----------------------------ProThrSerLysLysAspGlnValLysAsp 449

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              :::|||:::            ||||||      :::|||   :::   :::      :::
Db        450 AsnLysAsp---------LysGluLysAlaProLysGluGluMetLysSerValTrpAsp 466

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              :::|||   :::                                          |||

Db        467 ArgLysArgGlyVal----------------------------------------Pro 472

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||||||||           |||                 |||   |||      :::
Db        473 GluGlnLys-----------AlaGlnAsnGlyGluArgGluLeuThrThrProLysLeu 488

Qy       1000 ---TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
                 |||      |||  |||      |||      :::         :::|||||
Db        489 LysSerThrGluAsnAlaPheGlyArgSerAsnLeuLysGlyAlaAlaAsnAlaGluAla 508

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAA------GACTTAGACGAATTTAAG 1110
                        :::   |||::::::   :::         :::|||||||||   |||
Db        509 GlySerGluLysLeuLysGluLysGlnGlnGluAlaAlaValGluLeuAspGluLeuLys 528

Qy       1111 CCTATTGTGCAATATGACAAT------TTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                     :::                     ::::::||||||                :::
Db        529 LysArgArgGluGluArgArgLysIleLeuGluGluGluGluGlnLysLysLysGlnGlu 548

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
                 |||  |||      |||  :::  |||             :::  |||  ||||||
Db        549 GluAlaGluArgLysIleArgGluGluGluGluLysLysArgMetLysGluGluIleGlu 568

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA---------AAAGGAAAGAAATAT 1275
                 :::         |||  :::|||::::::  |||             |||
Db        569 ArgArgArgAlaGluAlaAlaGluLysArgGlnLysValProGluAspGlyValSerGlu 588

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATT 1335
                 ||||||              |||  |||        ||||||     |||
Db        589 GluLysLysProPheLysCysPheSerProLysGlySerSerLeuLysIleGluGluArg 608

Qy       1336 AAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
                 ::::::              :::||||||      :::|||
Db        609 AlaGluPhe--------------LeuAsnLysSerAlaGlnLys 618

RESULT 11
JC5421
smooth muscle myosin heavy chain 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C;Accession: JC5421
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5421
A;Molecule type: mRNA
A;Residues: 1-1938 <HAS>
A;Cross-references: UNIPROT:O08638; UNIPARC:UPI000002A11C; DDBJ:D85924; NID:g1945079; PI
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.:              5.23e-10        Length:         1938
Score:                  319.00          Matches:        144
Percent Similarity:     42.02%          Conservative:   114
Best Local Similarity:  23.45%          Mismatches:     180
Query Match:            12.83%          Indels:         176
DB:                     2               Gaps:           27

US-09-837-344-41 (1-1482) x JC5421 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||         ||||||
Db        932 GluGluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                            ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971
```

```
Qy       109 CAGCAA-------------------------------------------------AGCGAT 120
Db       972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
Db       992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

Qy       175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
Db      1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

Qy       196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
Db      1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy       250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
Db      1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy       286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
Db      1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy       304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG-------------------------- 336
Db      1092 LeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy       337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db      1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy       391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
Db      1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy       448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
Db      1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy       502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
Db      1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy       535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
Db      1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy       580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db      1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy       640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
Db      1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy       670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db      1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu---------------------- 1264

Qy       730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db      1265 ---------------------------------------------GlnSerLysCys 1268

Qy       790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
Db      1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy       850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db      1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

.                 ::::::      ||||||   :::            |||   :::|||
Db      1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db      1322 GluLeuGluGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db      1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db      1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db      1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
Db      1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy      1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------- 1290
Db      1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy      1291 ------------------------------------AATAATTTTAAACCAAATGAT 1311
Db      1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy      1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT---------------------- 1350
Db      1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy      1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
Db      1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy      1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
Db      1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
```

RESULT 12
JC5420
smooth muscle myosin heavy chain 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C;Accession: JC5420
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5420
A;Molecule type: mRNA
A;Residues: 1-1972 <HAS>
A;Cross-references: UNIPROT:O08638; UNIPARC:UPI00000279BF; DDBJ:D85923; NID:g1945077; PI
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:

Pred. No.:	5.21e-10	Length:	1972
Score:	319.00	Matches:	144
Percent Similarity:	42.02%	Conservative:	114
Best Local Similarity:	23.45%	Mismatches:	180
Query Match:	12.83%	Indels:	176
DB:	2	Gaps:	27

US-09-837-344-41 (1-1482) x JC5420 (1-1972)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||         ||||||
Db        932 GluGluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA----------------------------------------------AGCGAT 120
              ::::::                                              |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                           |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 -----------------------------------AGACTTGCTAAA------GAAAAG 303
                                                  :::|||||||||      |||:::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG-------------------------- 336
              ||||||         :::   |||::::::|||
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
                    |||      :::      :::   :::   |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::          :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::            |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                                  |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||:::      ::::::      :::||||||
Db       1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu-------------------- 1264

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                                 ||||||
Db       1265 ----------------------------------------------GlnSerLysCys 1268
```

```
Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                          |||      :::      |||   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                    ||||||:::|||                |||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ::::::      ||||||   :::            |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                    :::|||      |||   :::   :::|||:::|||                :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   ||||||||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::                |||         |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::            |||:::|||:::|||                |||   :::
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::   ::::::|||      |||   ||||||         ||||||
Db       1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------- 1290
              |||                |||   |||||||||||:::
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1291 ---------------------------------------AATAATTTTAAACCAAATGAT 1311
                                                       :::|||   :::
Db       1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT---------------------- 1350
              |||      :::|||:::   :::   :::   ||||||
Db       1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy       1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
                    |||:::         :::   :::||||||      |||:::|||      :::
Db       1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy       1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
                               |||   |||:::      :::||||||
Db       1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
```

RESULT 13
A61231
myosin heavy chain nonmuscle form A - human
N;Alternate names: cellular myosin heavy chain; myosin type 9; NMMHC-A
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 14-Jul-1994 #text_change 31-Dec-2004
C;Accession: A61231; A34876; I52562; I61692
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on differe
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: A61231
A;Molecule type: mRNA
A;Residues: 1-715 <SIM>
A;Cross-references: UNIPROT:P35579; UNIPROT:Q9UMJ0; UNIPARC:UPI000016ADA3; GB:M69180; NI
R;Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A;Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through alterna
A;Reference number: A34876; MUID:90138958; PMID:1967836
A;Accession: A34876

```
A;Molecule type: mRNA
A;Residues: 715-1961 <SAE>
A;Cross-references: UNIPARC:UPI000016ADA7; GB:M31013; NID:g189035; PIDN:AAA36349.1; PID:
R;Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.;
Blood 78, 1826-1833, 1991
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, c
A;Reference number: I52562; MUID:92003925; PMID:1912569
A;Accession: I52562
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-52,'EAI',56-659,'T',661-868,'T',870-930,'C',932-1239,'KG',1242-1337 <RES>
A;Cross-references: UNIPARC:UPI000016AD9A; GB:M81105; NID:g188988; PIDN:AAA59888.1; PID:
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A;Title: Identification and overlapping expression of multiple unconventional myosin gen
A;Reference number: A55758; MUID:94294418; PMID:8022818
A;Accession: I61692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 182-218 <BEM>
A;Cross-references: UNIPARC:UPI0000073258; GB:L29141; NID:g457249; PIDN:AAA20904.1; PID:
C;Genetics:
A;Gene: GDB:MYH9
A;Cross-references: GDB:120216; OMIM:160775
A;Map position: 22q12.3-22q13.1
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1938/Domain: coiled coil #status predicted <COI>
F;837-1277/Domain: S2 #status predicted <DS2>
F;1278-1961/Domain: light meromyosin #status predicted <LMM>
F;1939-1961/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694,704/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                5.54e-10      Length:          1961
Score:                    318.50        Matches:         145
Percent Similarity:       44.00%        Conservative:    97
Best Local Similarity:    26.36%        Mismatches:      171
Query Match:              12.81%        Indels:          137
DB:                       1             Gaps:            27

US-09-837-344-41 (1-1482) x A61231 (1-1961)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::     ||||||:::|||      |||:::::::|||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::         |||              ||||||   :::||||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
              ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
                               |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::|||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
              |||:::   :::||||||:::   ||||||         ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                 :::||||||      |||::::::|||      :::||||||   |||  :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||      :::   :::|||:::||||||         :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                               |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 -----------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                              ||||||   :::|||      |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                   ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   :::::::|||           |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                 ||||||||||||||
Db       1221 --------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT 834
                       |||   |||   |||   :::           |||::::::
Db       1232 GluValLysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy        835 ATA------------------ATAGAAAAAACAAATAGAGAATCTATTACAACA------ 870
                               ::::::           |||   :::   |||
Db       1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy        871 ------------------------------------AATGTTGAAGGA---------CGA 885
                                                  ||||||   |||        :::
Db       1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                 |||           |||           ||||||   |||           |||
Db       1292 SerAspSerLysSerSerLysLeu------ThrLysAspPheSerAlaLeuGluSerGln 1309

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              :::|||           :::|||           :::   :::::::::|||
Db       1310 LeuGlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr 1329

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                    ::::::   :::   |||||||||   ::::::   :::   :::
Db       1330 ------------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGln 1344

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||||||||:::   :::           |||   :::           :::
Db       1345 LeuGluGluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHis 1364

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                    |||           :::   |||::::::||||||   :::
Db       1365 AlaGlnValAlaAsp--------------MetLysLysLysMetGluAspSerValGly 1379

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
```

```
                  |||      :::::::::          ::::::|||      |||   ||||||
Db       1380 CysLeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSer 1399

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                     |||||||||:::           |||   ||||||||||
Db       1400 GlnArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeu 1419

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGAT 1353
                 :::      |||      |||   |||   :::|||      :::   ::::::
Db       1420 GlnGlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAla 1437

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
                    :::   |||:::::::::||||||
Db       1438 CysAsnLeuGluLysLysGlnLysLysPhe 1447

RESULT 14
H86341
hypothetical protein F9H16.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: H86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86341
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1166 <STO>
A;Cross-references: UNIPROT:Q9SYP6; UNIPARC:UPI00000A7C33; GB:AE005172; NID:g4836896; PI
C;Genetics:
A;Map position: 1

Alignment Scores:
Pred. No.:                6.82e-10        Length:          1166
Score:                    317.50          Matches:         133
Percent Similarity:       41.21%          Conservative:    106
Best Local Similarity:    22.93%          Mismatches:      172
Query Match:              12.77%          Indels:          169
DB:                       2               Gaps:            24

US-09-837-344-41 (1-1482) x H86341 (1-1166)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||:::::::::||||||::::::   ||||||   :::::::::      |||::::::
Db         17 GluGluGluAlaAspLeuLysLysSerArgArgAspArgAspArgSerAsnGluArgLys 36

Qy         64 AGCGAT-----------------TTAGAACAAGATAGACTTGCTAAA------------ 93
                 |||                     |||:::||||||      |||
Db         37 LysAspLysGlySerGluLysArgArgGluLysAspArgArgLysLysArgValLysSer 56

Qy         94 --------------------------------GAAAAGTTACAAGAGCAGCAAAGCGAT 120
                                              ::::::   :::|||::::::   :::
Db         57 SerAspSerGluAspAspTyrAspArgAspAspAspGluGluArgGluLysArgLysGlu 76

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAA------------------------------- 147
                 |||:::||||||      ::::::
Db         77 LysGluArgGluArgArgArgArgAspLysAspArgValLysArgArgSerGluArgArg 96

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG---------------AGACGTGCT 192
                 |||      :::   :::   |||:::|||:::|||               ||||||
Db         97 LysSerSerAspSerGluAspAspValGluGluGluAspGluArgAspLysArgArgVal 116

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--------------- 237
                 ||||||          ::::::          |||   ||||||
Db        117 AsnGluLys---------GluArgGlyHisArgGluHisGluArgAspArgGlyLysAsp 133

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
              |||   :::::::::   :::|||::::::   |||   |||:::::::|||
Db        134 ArgLysArgAspArgGluArgGluGluArgLysAspLysGluArgGluArgGluLysAsp 153

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
                 :::   :::|||::::::   :::       ::::::   |||   ||||||:::
Db        154 ArgGluArgArgGluArgGluArgGluGluArgGluLysGluArgValLysGluArgGlu 173

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
                 :::   ::::::   |||   |||::::::||||||   :::||||||   :::
Db        174 ArgArgGluArgGluAspGlyGluArgAspArgArgGluArgGluLysGluArgGlySer 193

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
                 :::   :::   |||:::||||||       :::          :::|||          |||
Db        194 ArgArgAsnArgGluArgGluArgSerArgGluValGlyAsnGluGluSerAspAspAsp 213

Qy        478 TTAGAACAAGAG------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531
                 ::::::::::::      ||||||   ||||||       :::   ::::::   :::
Db        214 ValLysArgAspLeuLysArgArgArgLysGluGlyGlyGluArgLysGluLysGluArg 233

Qy        532 GAACAA----------------------GAGAGACGTGCTAAAGAAAAGTTGCAAGAA 567
                 |||:::                        |||          |||   |||      |||
Db        234 GluLysSerValGlyArgSerSerArgHisGluAspSerProLysArgLysSerValGlu 253

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                    :::   ::::::::::::      :::|||:::|||:::::::::|||:::
Db        254 AspAsnGlyGluLysLysGluLysLysThrArgGluGluGluLeuGluAspGluGlnLys 273

Qy        628 GATTTAGAACAA-----------------AGGAAGGCTGATACGAAAAAAAATTTAGAA 669
                 |||::::::                |||:::          :::   |||:::
Db        274 LysLeuAspGluGluValGluLysArgArgArgArgValGlnGluTrpGlnGluLeuLys 293

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||||||||
Db        294 ArgLysLysGlu---------------------------------------------- 297

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    |||      ||||||::::::
Db        298 ------GluAlaGluSerGluSerLys-------------------------------- 304

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
                    |||      ||||||          :::   :::
Db        305 ---GlyAspAlaAspGlyAsnGluProLysAlaGlyLys-------------------- 316

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                                          :::||||||      |||      :::||||||
Db        317 --------------AlaTrpThrLeuGluGlyGluSerAspAspGluGluGlyHisPro 331

Qy        910 GAAGAAAAGAAA-----------------GATGGTTCAATAAAACCAGAACAAAAAGAA 951
                 ||||||||                     |||         ||||||||
Db        332 GluGluLysSerGluThrGluMetAspValAspGluGluThrLysProGluAsnAspGly 351

Qy        952 GAT---AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
                 |||   |||      |||::::::|||   |||      ||||||:::   |||
Db        352 AspAlaLysMetValAspLeuGluAsnGluThrAlaAlaThrValSerGluSerGly--- 370

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                                                                    |||
Db        371 ----------------------------------------------GlyAspGlyAla 374

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT---------------------- 1107
                 :::||||||||||||   |||      ||||||   |||
Db        375 ValAspGluGluGluIleAsp---ProLeuAspAlaPheMetAsnThrMetValLeuPro 393

Qy       1108 --------------------------AAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                                         |||          |||          |||
Db        394 GluValGluLysPheCysAsnGlyAlaProProProAlaValAsnAspGlyThrLeuAsp 413

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA------AATGAA 1194
                 :::      |||      ||||||      |||      :::|||      |||:::
Db        414 SerLysMetAsnGly------LysGluSerGlyAspArgProLysLysGlyPheAsnLys 431
```

```
Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
Db        432 AlaLeuGlyArgIleIleGlnGly---GluAspSerAspSerAspTyrSerGlu------ 448

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
Db        449 --------------------ProLysAsnAspAspAspProSerLeuAspGluAspAsp 461

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG---------------CAG 1359
Db        462 GluGluPheMetLysArgValLysLysThrLysAlaGluLysLeuSerLeuValAspHis 481

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
Db        482 SerLysIleGluTyrGluProPheArgLysAsn---PheTyrIle--------------- 495

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
Db        496 GluValLysAspIleSerArgMetThrGlnGluGluValAsnThrTyrArgLysGluLeu 515

RESULT 15
I38055
myosin heavy chain, perinatal skeletal muscle - human
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jul-2004
C;Accession: I38055; JH0154; S12459; S09332; A30220; S49478
R;Jullian, E.H.; Kelly, A.M.; Pompidou, A.J.; Hoffman, R.; Schiaffino, S.; Stedman, H.H.
Eur. J. Biochem. 230, 1001-1006, 1995
A;Title: Characterization of a human perinatal myosin heavy-chain transcript.
A;Reference number: I38055; MUID:95324556; PMID:7601129
A;Accession: I38055
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1937 <RES>
A;Cross-references: UNIPROT:P13535; UNIPARC:UPI000016ADA9; EMBL:Z38133; NID:g558668; PID
R;Karsch-Mizrachi, I.; Feghali, R.; Shows, T.B.; Leinwand, L.A.
Gene 89, 289-294, 1990
A;Title: Generation of a full-length human perinatal myosin heavy-chain-encoding cDNA.
A;Reference number: JH0154; MUID:90323631; PMID:2373371
A;Accession: JH0154
A;Molecule type: mRNA
A;Residues: 1-14,'A',16-859 <KAR>
A;Cross-references: UNIPARC:UPI000017760F; GB:Y00821
A;Experimental source: skeletal muscle
R;Bober, E.
submitted to the EMBL Data Library, January 1989
A;Reference number: S12458
A;Accession: S12459
A;Molecule type: mRNA
A;Residues: 502-1071,'N',1073-1250,'DGG',1253-1376,'NT',1379-1913,'D',1915-1937 <BOB>
A;Cross-references: UNIPARC:UPI000016A5F8; EMBL:X51592; NID:g29465; PIDN:CAA35941.1; PID
A;Experimental source: clone gtMHC-F
R;Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
Eur. J. Biochem. 189, 55-65, 1990
A;Title: Identification of three developmentally controlled isoforms of human myosin hea
A;Reference number: S09331; MUID:90235862; PMID:1691980
A;Accession: S09332
A;Molecule type: mRNA
A;Residues: 502-547,'X',549-617,'X',619-687,'X',689-757,'X',759-827,'X',829-897,'X',899-
-1376,'NT',1379-1386,'X',1388-1456,'X',1458-1526,'X',1528-1596,'X',1598-1666,'X',1668-17
A;Cross-references: UNIPARC:UPI0000177610; EMBL:X51592
R;Feghali, R.; Leinwand, L.A.
J. Cell Biol. 108, 1791-1797, 1989
A;Title: Molecular genetic characterization of a developmentally regulated human perinat
A;Reference number: A30220; MUID:89234168; PMID:2715179
A;Accession: A30220
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 860-969,'Q',971-1246,'H',1248-1260,'G',1262-1296,'Q',1298-1503,'AH',1506-184
A;Cross-references: UNIPARC:UPI0000177611; GB:Y00821; NID:g34863; PIDN:CAA68757.1; PID:g
C;Genetics:
A;Gene: GDB:MYH8
A;Cross-references: GDB:125267; OMIM:160741
A;Map position: 17pter-17p12
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b
F;91-769/Domain: myosin motor domain homology <MMOT>
F;181-188/Region: nucleotide-binding motif A (P-loop)
F;551-588/Region: actin binding #status predicted
F;658-680/Region: actin binding #status predicted
F;842-1282/Region: S2 #status predicted
F;698,708/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                6.26e-10      Length:          1937
Score:                    317.50        Matches:         156
Percent Similarity:       42.08%        Conservative:    131
Best Local Similarity:    22.87%        Mismatches:      184
Query Match:              12.77%        Indels:          211
DB:                       2             Gaps:            35

US-09-837-344-41 (1-1482) x I38055 (1-1937)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189
```

```
Qy       540 ------------------------------------------------------------ 540

Db      1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy       541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                         :::|||   |||:::||||||:::::::::::|||:::|||:::
Db      1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy       589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
             |||                                :::||||||     |||:::
Db      1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy       616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
             :::      |||   ||||||           |||          |||:::   |||
Db      1250 LysMetCysArgSerLeuGlu-----------AspGlnValSerGluLeuLysThrLys 1265

Qy       676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
             :::|||            |||      ||||||               ||||||:::
Db      1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy       727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
                |||   |||          :::   :::|||:::         :::         |||
Db      1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy       784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATA------TCTATA 837
                |||   :::      :::   :::::::::      :::|||:::         :::
Db      1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy       838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
             |||:::|||      :::   ::::::          ::::::      |||
Db      1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy       889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
             |||:::   :::          ||||||:::::::::|||
Db      1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy       933 ------------------------------------------------------------ 933

Db      1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy       934 --------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                          :::   |||:::   :::   |||   |||
Db      1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy       979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
                |||   |||          :::                 :::   |||   :::
Db      1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy      1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
             :::      |||   :::|||   :::::::|||   |||
Db      1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy      1087 -----------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                          |||:::   |||   |||:::|||
Db      1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy      1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------------- 1152
                :::|||:::      |||   |||
Db      1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy      1153 ------------------------GGAATTTATAAAGAA------------CTAGAAGAT 1176
                                       :::|||:::|||            ||||||
Db      1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy      1177 TTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGATGAAGGAATAGAA 1224
             |||      :::|||:::||||||               ||||||   |||   |||
Db      1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy      1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
             :::      :::::::::   |||   ||||||||||||||      :::   ||||||
Db      1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy      1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
```

```
                :::               :::      :::   ||||||      |||      |||
Db      1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy      1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
                   :::          :::::|||:::   |||   :::      :::
Db      1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy      1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
                |||   :::   |||   ||||||           |||:::|||      :::
Db      1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy      1459 ACTAAA 1464
             |||:::
Db      1594 ThrArg 1595

RESULT 16
A47297
myosin heavy chain form B, nonmuscle - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47297; A55441
R;Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.
Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993
A;Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscl
A;Reference number: A47297; MUID:93219383; PMID:8464900
A;Accession: A47297
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1992 <BHA>
A;Cross-references: UNIPROT:Q04834; UNIPARC:UPI00000FC054; GB:L09740; NID:g214623; PIDN:.
A;Experimental source: XTC cells
A;Note: sequence extracted from NCBI backbone (NCBIP:128722)
R;Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.
J. Biol. Chem. 270, 1395-1401, 1995
A;Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34(
A;Reference number: A55441; MUID:95138137; PMID:7836406
A;Accession: A55441
A;Status: preliminary
A;Molecule type: protein
A;Residues: 198-232 <KEL>
A;Cross-references: UNIPARC:UPI0000177615
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-787/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.:              6.23e-10        Length:         1992
Score:                  317.50          Matches:        133
Percent Similarity:     45.94%          Conservative:   116
Best Local Similarity:  24.54%          Mismatches:     171
Query Match:            12.77%          Indels:         122
DB:                     2               Gaps:           23

US-09-837-344-41 (1-1482) x A47297 (1-1992)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAA---------GAGAGACGTGCTAAAGAAAAGTTG 51
              |||             |||||||||:::         |||      |||   :::||||||
Db        965 GlnThrHisValGlnAspLeuGluGluGlnLeuAspGluGluGluAlaAlaGlnLysLeu 984

Qy         52 CAAGAACAACAA--------------------AGCGATTTAGAACAAGATAGACTTGCT 90
              |||   ::::::                        :::|||:::|||   |||
Db        985 GlnLeuGluLysValThrAlaGluAlaLysIleLysLysMetGluGluAspIleLeuVal 1004

Qy         91 AAAGAAAAGTTACAAGAGCAGCAAAGCGAT---TTAGAACAAGAGAGACTTGCTAAAGAA 147
                    |||::::::|||   |||      |||:::::::::::::|||   :::|||
Db        1005 ---------LeuGluAspGlnAsnSerLysPheLeuLysGluLysLysLeuLeuGluGlu 1021

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAA------ 198
              ::::::   |||      |||   |||   |||:::|||   :::||||||
Db        1022 ArgIleAlaGluSerThrSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAla 1041

Qy        199 AAGTTGCAAGAACAACAA------------AGCGATTTAGAACAAGAG------------ 234
```

```
Db       1042 LysLeuLysAsnLysGlnGluMetMetIleSerAspLeuGluGluArgLeuLysLysGlu 1061

Qy        235 --------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273

Db       1062 GluLysThrArgGlnGluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAsp 1081

Qy        274 TTAGAACAAGATAGACTTGCTAAA--------------GAAAAGTTACAAGAGCAGCAA 318

Db       1082 Phe---GlnAspGlnIleAlaGluLeuGlnAlaGlnIleGluGluLeuLysLeuGlnLeu 1100

Qy        319 AGCGATTTAGAACAAGAG--------------AGACGTGCTAAAGAAAAGTTGCAAGAA 363

Db       1101 AlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgGlyAspGluGluValLeuGln 1120

Qy        364 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 423

Db       1121 LysAsnAsnThrLeuLysLeuValArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGlu 1140

Qy        424 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 483

Db       1141 AspLeuGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSer 1160

Qy        484 CAAGAGAGACGTGCT---------------------------------------------- 498

Db       1161 GluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGln 1180

Qy        499 --------------------AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537

Db       1181 GlnGluLeuArgThrLysArgGluGlnGluValAlaGluLeuArgLysSerIleGluGlu 1200

Qy        538 GAGAGACGTGCTAAAGAA----------------------------------AAGTTG 561

Db       1201 GluThrArgAsnHisGluAlaGlnIleGlnGluMetArgGlnArgGlnAlaThrAlaLeu 1220

Qy        562 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621

Db       1221 GluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLysValAsnLeuGluLysAsn 1240

Qy        622 CAAAGAGATTTAGAA---------CAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672

Db       1241 LysGlnSerLeuGluSerAspAsnLysGluLeuAlaThrGluValLysSerLeuGlnGln 1260

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA---GAAATACCA 729

Db       1261 MetLysAlaGluSerGluTyrLysArgLysLysLeuGluGlyGlnValGlnGluLeuHis 1280

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789

Db       1281 AlaLysValLeuGluGlyAspArgLeuArgAlaAspMetValGluLysSerSerLysLeu 1300

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849

Db       1301 GlnAsnGluLeuGluAsnValSer----------------SerLeuLeuGluGluAla 1314

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909

Db       1315 GluLysLysGlyIle------------------------------------------- 1319

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969

Db       1320 LysLeuAlaLysAspValAlaSerMetGluSerGlnLeuGlnAspThrGlnGluLeuLeu 1339

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029

Db       1340 GlnGluGluThrArgGlnLysLeuAsnGlnSerSer------------ArgIleArgGln 1355

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089

Db       1356 LeuGluGluGlu--------------LysAsnAsnLeuGlnGluGlnGlnGluGluGlu 1370

Qy       1090 GAAGACTTAGACGAATTT------AAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143

Db       1371 GluGluAlaArgLysSerLeuGluLysGlnIleLeuSerLeuGlnSer------------ 1386

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203

Db       1387 GlnLeuIleGluAlaLysLysLysValAspAspGluValGlyThrIleGluGlyLeuGlu 1406

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAAAAA 1254

Db       1407 GluValLysLysLysLeuLeuLysAspThrGluGlyLeuGlyGlnArgLeuGluGluLys 1426

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314

Db       1427 IleIleAlaTyrGluLysLeuGluLysThrLys---AsnArg-LeuGlnGlnGluLeuAs 1445

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374

Db       1445 pAspLeuMetValAspLeuAspHisGlnArgGlnIleValSerAsnLeuGluLysLysGl 1465

Qy       1375 GAAA 1378

Db       1465 nLys 1466
```

RESULT 17
B85431
trichohyalin like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: B85431
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: B85431
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1432 <STO>
A;Cross-references: UNIPROT:O23230; UNIPARC:UPI000009D2F0; GB:NC_001268; NID:g7270600; P
C;Genetics:
A;Gene: AT4g36520
A;Map position: 4

Alignment Scores:

Pred. No.:	7e-10	Length:	1432
Score:	317.00	Matches:	136
Percent Similarity:	43.26%	Conservative:	108
Best Local Similarity:	24.11%	Mismatches:	196
Query Match:	12.75%	Indels:	124
DB:	2	Gaps:	22

US-09-837-344-41 (1-1482) x B85431 (1-1432)

```
Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78

Db        632 MetGluGluGluThrArgIleLysGluAlaArgLeuArgGluGluAsnAsp---ArgArg 650

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAG---------CAAAGCGATTTAGAACAA 129

Db        651 GluArgValAlaValGluLysAlaGluAsnGluLysArgLeuLysAlaAlaLeuGluGln 670

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189

Db        671 GluGlu---LysGluArgLysIleLysGluAlaArgGluLysAlaGluAsnGluArgArg 689

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249

Db        690 AlaValGluAlaArgGluLys------------AlaGluGlnGluArgLysMetLysGlu 705

Qy        250 -----------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291

Db        706 GlnGlnGluLeuGluLeuGlnLeuLysGluAlaPheGluLysGluGluGluAsnArgArg 725

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
```

```
Db        726 MetArgGluAlaPheAlaLeuGluGln---------GluLysGluArgArgIleLysGlu 742

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTG 408
              :::::::::::::::::         :::|||   |||   |||   :::|||
Db        743 AlaArgGluLysGluGluAsnGluArgArgIleLysGluAlaArgGluLysAlaGluLeu 762

Qy        409 CAAGAACAA---CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::::::::   ::::::  ||||||||||||         :::   :::::::::|||
Db        763 GluGlnArgLeuLysAlaThrLeuGluGlnGluGlu---LysGluArgGlnIleLysGlu 781

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
              :::|||         |||   ||||||||||||||||||   |||::::::
Db        782 ArgGlnGluArgGluGluAsnGluArgArgAlaLysGluValLeuGluGln--------- 798

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                  |||   ||||||:::   ||||||   |||::::::::::           |||
Db        799 ---AlaGluAsnGluArgLysLeuLysGluAlaLeuGluGlnLys-----------Glu 813

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
              |||||||||   ||||||      ::::::::::          :::   |||:::
Db        814 AsnGluArgArgLeuLysGluThrArgGluLysGluGluAsnLysLysLysLeuArgGlu 833

Qy        646 GCT-----------GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATA 693
              |||           :::   :::|||   |||             |||      :::|||
Db        834 AlaIleGluLeuGluGluLysGluLysArgLeuIleGluAlaPheGluArgAlaGluIle 853

Qy        694 ---------TTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCA 744
                      |||   |||||||||                       ::::::
Db        854 GluArgArgLeuLysGluAspLeuGluGlnGluGluMetArgMetArgLeuGlnGluAla 873

Qy        745 TCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA 804
              |||   ||||||               |||:::   :::      |||:::::::::
Db        874 LysGluArgGluArg-----------LeuHisArgGluAsn------GlnGluHisGln 887

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATT 864
              |||   |||          |||   |||      |||          :::|||
Db        888 GluAsnGluArgLysGlnHisGluTyrSerGlyGluGluSerAspGluLysGluArgAsp 907

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAA---GGACATCTTGAAGAAAAGAAA 921
                  :::|||         :::      |||      |||   |||:::
Db        908 AlaCysGluMetGluLysThrCysGluThrThrLysGluAlaHisGlyGluGlnSerSer 927

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
              :::   |||:::           :::|||::::::|||   |||
Db        928 AsnGluSerLeuSerAspThrLeuGluGluAsnGluSerIleAsp--------------- 942

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
                                  ||||||      ::::::|||   :::   |||
Db        943 --------------------------AsnAspValSerValAsnLysGlnLysLysGlu 953

Qy       1042 --------------------------ATAAGTGCTGAA---------------------- 1053
                                       :::|||||||||
Db        954 GluGluGlyThrArgGlnArgGluSerMetSerAlaGluThrCysProTrpLysValPhe 973

Qy       1054 ------------------------------------------------TATGACGAT 1062
                                                             |||
Db        974 GluLysThrLeuLysAspAlaSerGlnLysGluGlyThrAsnGluMetAspAlaAspThr 993

Qy       1063 TCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA 1122
              |||   :::      |||:::         |||   |||                  |||
Db        994 ArgLeuPheGluArgAsnGluGluThrProArgLeuGlyGluAsnGlyGlyCysAsnGln 1013

Qy       1123 TATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA--------------GAA 1167
                 :::                |||||:::      :::   :::             :::
Db       1014 GlnAsnGlyGluSerGlyGluGluSerThrSerValThrGluAsnIleIleGlyGlyLys 1033

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT-------------- 1212
              ||||||         :::   :::|||         |||
Db       1034 LeuGluGlnLysSerLysAsnSerGluThrSerLysAspAlaSerValLeuLysArgVal 1053

Qy       1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAA 1272
              |||::::::         ||||||      |||:::   :::      |||:::
Db       1054 SerGlyLeuLysThrGluValGluGluArgLeuGluAspValValGlyValGlyArgAsp 1073

Qy       1273 TATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332
                          :::::::::         |||                  :::      |||
Db       1074 GlnArgAsnProGluGluSerLysSerAlaProLysThrSerTyrGlyPheArgAsnHis 1093

Qy       1333 ATTAAAAAATATAAAAATGATAAGCAG-------------------------------- 1359
                    |||:::   :::   ::::::
Db       1094 GluTyrLysPheThrHisGlnGlnGluArgGlyAsnIleTyrGluThrGlnAlaGlyLeu 1113

Qy       1360 -------------------------------------GTTAATAAGGAAAAGGAAAAA 1380
                                                   |||:::   ::::::||||||
Db       1114 AsnGlnAspAlaLysValGluArgProLeuProSerArgValSerValGlnArgGluLys 1133

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                    :::   |||             |||   :::   |||   |||:::   |||:::
Db       1134 GluAlaGluArgLeuLysArgGluArgAspLeuGluMetGluGlnLeuArgLysValGlu 1153

Qy       1441 GAGTTATCTGAA 1452
              |||      |||
Db       1154 GluGluArgGlu 1157
```

RESULT 18
A29320
myosin heavy chain, fast skeletal muscle, embryonic [similarity] - chicken
N;Alternate names: myosin heavy chain, EFW1
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C;Accession: A29320; B24124; A02990
R;Molina, M.I.; Kropp, K.E.; Gulick, J.; Robbins, J.
J. Biol. Chem. 262, 6478-6488, 1987
A;Title: The sequence of an embryonic myosin heavy chain gene and isolation of its corre
A;Reference number: A29320; MUID:87194881; PMID:3571266
A;Accession: A29320
A;Molecule type: DNA; mRNA
A;Residues: 1-1940 <MOL>
A;Cross-references: UNIPROT:P02565; UNIPARC:UPI00001775FE
A;Note: the sequence translated from the mRNA differs from that of the DNA in having 379
R;Gulick, J.; Kropp, K.; Robbins, J.
J. Biol. Chem. 260, 14513-14520, 1985
A;Title: The structure of two fast-white myosin heavy chain promoters. A comparative stu
A;Reference number: A92507; MUID:86033956; PMID:2997212
A;Accession: B24124
A;Molecule type: DNA
A;Residues: 1-168 <GUL>
A;Cross-references: UNIPARC:UPI00001775FF
R;Kavinsky, C.J.; Umeda, P.K.; Sinha, A.M.; Elzinga, M.; Tong, S.W.; Zak, R.; Jakovcic,
J. Biol. Chem. 258, 5196-5205, 1983
A;Title: Cloned mRNA sequences for two types of embryonic myosin heavy chains from chick
A;Reference number: A02990; MUID:83161144; PMID:6833296
A;Accession: A02990
A;Molecule type: mRNA
A;Residues: 1502-1546,'A',1548-1912,'G',1914,'T',1916-1940 <KAV>
A;Cross-references: UNIPARC:UPI000017136C; GB:V00430; GB:J00892; NID:g63619; PIDN:CAA237
A;Experimental source: clone 251
C;Comment: The entire sequence of this fragment corresponds to the light meromyosin port
ides (heptads I-IV), that are characteristic of coiled coils.
C;Comment: Association of the alpha-helical regions of two myosin heavy chains to form a
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; methylated amino acid; muscle; muscle contr
F;89-771/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;130/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted

Alignment Scores:

Pred. No.:	7.5e-10	Length:	1940
Score:	316.00	Matches:	163
Percent Similarity:	41.27%	Conservative:	123
Best Local Similarity:	23.52%	Mismatches:	181
Query Match:	12.71%	Indels:	226
DB:	2	Gaps:	38

US-09-837-344-41 (1-1482) x A29320 (1-1940)

```
Qy        4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db      932 GluAspGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 951

Qy       61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
Db      952 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 971

Qy       94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db      972 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaAlaLeuAsp 991

Qy      127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db      992 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1011

Qy      169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db     1012 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1031

Qy      202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db     1032 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1051

Qy      238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db     1052 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysMetThrGlnGluSerThr 1071

Qy      268 AGCGATTTAGAACAAGATAGA---CTTGCTAAAGAAAAGTTACAAGAGCAG--------- 315
Db     1072 MetAspLeuGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysAspPheGlu 1091

Qy      316 ------------CAAAGCGATTTAGAACAAGAG----------------------AGA 339
Db     1092 IleSerGlnIleGlnSerLysIleGluAspGluGlnAlaLeuGlyMetGlnLeuGlnLys 1111

Qy      340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db     1112 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluIleGluAlaGluArg 1131

Qy      391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db     1132 ThrSerArgAlaLysAlaGluLysHisArgAlaAspLeuSerArgGluLeuGluGluIle 1151

Qy      448 AAAGAAAAGTTGCAAGAA------------------CAACAAAGCGATTTAGAACAAGAG 489
Db     1152 SerGluArgLeuGluGluAlaGlyGlyAlaThrAlaAlaGlnIleAspMetAsnLysLys 1171

Qy      490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------- 540
Db     1172 ArgGluAla------GluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGln 1189

Qy      540 ------------------------------------------------------------ 540
Db     1190 HisGluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerThrAlaAspValGly 1209

Qy      541 ------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
Db     1210 GluGlnIleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeu 1229

Qy      583 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 609
Db     1230 LysMetGluIleAspAspLeuAlaSerAsnMetGluSerValSerLysAlaLysAlaAsn 1249

Qy      610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTA 666
Db     1250 LeuGluLysMetCysArgSerLeuGluAspGlnLeuSerGluIleLysThrLysGluGlu 1269

Qy      667 GAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATA 726
Db     1270 GluGlnGlnArgThrIleAsnAspIleSerAlaGlnLys------AlaArgLeuGlnThr 1287

Qy      727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786

.
Db     1288 GluSerGlyGluTyrSerArgGlnValGluGlu--------------LysAspAlaLeu 1302

Qy      787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAA 846
Db     1303 IleSerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLysArg 1322

Qy      847 ACAAATAGAGAATCTATTACAACAAAT--------------------GTTGAAGGACGA 885
Db     1323 HisLeuGluGluGluIleLysAlaLysLysCysProAlaHisAlaLeuGlnSerAlaArg 1342

Qy      886 AGGGATATACAT-----------------------------------AAAGGACATCTT 909
Db     1343 HisAspCysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluAlaLysGlyGluLeu 1362

Qy      910 GAAGAA-------------------------------------------------AAGAAA 921
Db     1363 GlnArgAlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThr 1382

Qy      922 GATGGTTCAATAAAACCAGAACAAAAAGAAGAT--------------------AAATCT 960
Db     1383 AspAlaIleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeu 1402

Qy      961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db     1403 GlnAspAlaGluGluHis---ValGluAlaValAsnSerLysCysAlaSer--------- 1418

Qy     1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db     1419 LeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGlu 1438

Qy     1081 GAAGAT-----------------------------------------GATGAAGACTTA 1098
Db     1439 ArgSerAsnAlaAlaCysAlaAlaLeuAspLysLysGlnLysAsnPheAspLysIleLeu 1458

Qy     1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA--------------- 1143
Db     1459 SerGluTrpLysGln------LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGln 1476

Qy     1144 --------------------------------GAAAACATAGGAATTTATAAAGAA--- 1167
Db     1477 LysGluSerArgSerLeuSerThrGluLeuPheLysMetLysAsnAlaTyrGluGluSer 1496

Qy     1168 ---------CTAGAAGATTTGATAGAGAAAAATGAAAATTTA-----------GATGAT 1206
Db     1497 LeuAspHisLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAsp 1516

Qy     1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAA 1263
Db     1517 LeuThrGluGlnIleAlaGluGlyGlyLysAlaIleHisGluLeuGluLysValLysLys 1536

Qy     1264 GGAAAGAAATATGAAAAAACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
Db     1537 GlnIleGluGlnGluLysSerGluLeuGlnThrAlaLeuGluGluAlaGluAlaSerLeu 1556

Qy     1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db     1557 GluHisGluGluGlyLysIleLeuArgValGlnLeuGluLeuAsnGlnValLysSerAsp 1576

Qy     1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT--------------- 1425
Db     1577 IleAspArgLys---------IleAlaGluLysAspGluGluIleAspGlnLeuLysArg 1593

Qy     1426 ------TTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
Db     1594 AsnHisLeuArgValValAspSerMetGlnSerThrLeu 1606

RESULT 19
A59293
skeletal myosin heavy chain - domestic rabbit
C;Species: Oryctolagus cuniculus
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: A59293
```

```
R;Maeda, K.; Hostinova, E.; Roesc.Kleinkauf, A.; Schuster, H.; Gasperik, J.; Wittinghofe
submitted to GenBank, July 1995
A;Description: Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal mus
A;Reference number: A59293
A;Accession: A59293
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1938 <MAE>
A;Cross-references: UNIPROT:Q28641; UNIPARC:UPI000012FB71; GB:U32574; NID:g940232; PIDN:
A;Experimental source: strain New Zealand White; cell type skeletal muscle fiber type II
C;Genetics:
A;Gene: MHC
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;89-769/Domain: myosin motor domain homology <MMO>

Alignment Scores:
Pred. No.:              9.55e-10        Length:         1938
Score:                  314.00          Matches:        152
Percent Similarity:     42.32%          Conservative:   129
Best Local Similarity:  22.89%          Mismatches:     179
Query Match:            12.63%          Indels:         204
DB:                     2               Gaps:           36

US-09-837-344-41 (1-1482) x A59293 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   |||||:::::::::
Db        930 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                ||||:::|||::::::|||                          ::::::   |||
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::          :::|||   |||||||||      |||||||||      ||||||:::
Db        990 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                     :::||||||   |||
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGACGTGCTAAAGAA 249
              |||::::::|||     |||             ||||||||||||::::::   :::
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysIleArgMet 1049

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
                |||:::       :::      ||||||   |||   :::||||||:::|||
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        307 CAAGAGCAGCAAAGCGAT---------TTAGAACAAGAGAGACGTGCTAAAGAA------ 351
                :::   ::::::|||         |||:::::::::   :::   ||||||
Db       1070 MetAspIleGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysGluPheGlu 1089

Qy        352 ------------------AAGTTGCAAGAACAACAAAGC---GATTTAGAACAAGAGAGA 390
                               |||:::::::::::|||:::      ::::::   ::::::
Db       1090 MetSerAsnLeuGlnSerLysIleGluAspGluGlnAlaLeuAlaMetGlnLeuGlnLys 1109

Qy        391 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
              :::   ||||||         :::::::::|||   :::   ::::::|||   ||||||
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1129

Qy        442 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACGTGCT 498
                ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        499 AAAGAAAAGTTGCAAGAA-----------CAACAAAGCGATTTAGAACAAGAGAGACGT 546
                |||:::|||:::|||                :::   :::|||      ::::::
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1169

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---------- 591
                 :::   :::   |||:::   :::   |||||||||:::
Db       1170 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        591 ---------------------------------------------------------- 591

Db       1190 AlaThrAlaAlaThrLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1209

Qy        592 ------------AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                        :::|||   |||:::||||||:::::::::::   :::|||:::
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACATGGAGATATATTA 696
                  |||         |||:::|||         |||      |||::::::
Db       1230 GluIleAspAspLeuAlaSerAsnMetGluThrValSerLysAlaLysGlyAsnLeu--- 1248

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756
                 |||   :::         ||||||            ||||||   ::::::   |||
Db       1249 ---GluLysMetCysArgThrLeuGluAspGlnValSerGluLeuLysThrLysGluGlu 1267

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
                                 ||||||      :::         :::         :::|||
Db       1268 Glu---------------HisGlnArgLeuIleAsnAspLeuSerAlaGlnArgAlaArg 1282

Qy        817 ---------GATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
                        :::|||   |||   |||      :::         :::::::|||:::
Db       1283 LeuGlnThrGluSerGlyGluPheSer---ArgGlnLeuAspGluLysAspSerLeuVal 1301

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
                 :::   :::         :::                  :::||||||   |||
Db       1302 SerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLys------ 1319

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
                  :::   ::::::|||   |||:::                  |||   |||:::
Db       1320 ------ArgGlnLeuGluGluGluIleLysAlaLysSerAlaLeuAlaHisAlaLeuGln 1337

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
                  :::                  |||           :::||||||:::|||
Db       1338 SerAlaArgHisAsp------CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGlu 1355

Qy       1048 GCT---------------------------------------------------------- 1050
              |||
Db       1356 AlaLysAlaGluLeuGlnArgAlaMetSerLysAlaAsnSerGluValAlaGlnTrpArg 1375

Qy       1051 ---GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                  :::|||:::          |||         |||:::   |||:::
Db       1376 ThrLysTyrGluThrAspAlaIleGlnArgThrGluGluLeuGluGluAlaLys------ 1393

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGAT---GAAGAAAACATA------------ 1152
              |||   :::   |||           ||||||   |||||::::::
Db       1394 LysLysLeuAlaGln------ArgLeuGlnAspAlaGluGluHisValGluAlaValAsn 1411

Qy       1153 --------------------------------GGAATTTATAAAGAACTAGAAGATTTG 1179
                                             :::      |||:::||||||||||
Db       1412 AlaLysCysAlaSerLeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeu 1431

Qy       1180 ATA---------------------------------------GAGAAAAATGAAAAT 1197
              :::                                     :::|||      |||
Db       1432 MetIleAspValGluArgThrAsnAlaAlaCysAlaAlaLeuAspLysLysGlnArgAsn 1451

Qy       1198 TTAGATGATTTA---------------------------------------GATGAA 1215
                 |||   :::                                       |||
Db       1452 PheAspLysIleLeuAlaGluTrpLysHisLysTyrGluGluThrHisAlaGluLeuGlu 1471

Qy       1216 GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
                ::|||   |||      ||||||   |||      |||               |||
Db       1472 AlaSerGlnLysGluSerArgSerLeuSerThrGluValPheLysValLysAsnAlaTyr 1491

Qy       1276 GAAAAAACAAAGGAT------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
                |||::::::   |||           |||      :::|||:::|||      |||
Db       1492 GluGluSerLeuAspGlnLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGlu 1511

Qy       1330 ---------------------------------------------CATATTAAAAAA 1341
```

```
                                                                  ::::::|||
Db       1512 IleSerAspLeuThrGluGlnIleAlaGluGlyGlyLysArgIleHisGluLeuGluLys 1531

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAG--------------------GAAAAA 1380
                 |||      |||||||||   :::||||||                    |||:::
Db       1532 ValLys------LysGlnValGluGlnGluLysSerGluLeuGlnAlaAlaLeuGluGlu 1549

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                    ||||||   |||           :::   :::||||||:::|||
Db       1550 AlaGluAlaSerLeuGluHis---------GluGluGlyLysIleLeuArgIleGlnLeu 1566

Qy       1441 GAGTTATCTGAA 1452
              ||||||::::::
Db       1567 GluLeuAsnGln 1570

RESULT 20
A59252
myosin heavy chain, nonmuscle, form IIB - human
N;Alternate names: myosin type 10; NMMHC-B
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A59252; B61231; G02055
R;Phillips, C.L.; Yamakawa, K.; Adelstein, R.S.
J. Muscle Res. Cell. Motil. 16, 379-389, 1995
A;Title: Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis
A;Reference number: A59252; MUID:96025307; PMID:7499478
A;Accession: A59252
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1976 <SIM>
A;Cross-references: UNIPROT:P35580; UNIPARC:UPI000012FB83; GB:M69181; NID:g641957; PIDN:
A;Experimental source: clone lib Lambda Zap II adult human T-cell library; cell line Jur
A;Note: between nucleotides 1942-1943 in mRNA encoding human brain MHC-B there is an alt
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on differe
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: B61231
A;Molecule type: mRNA
A;Residues: 63-237,'K',239-664,'L',666-722 <SI2>
A;Cross-references: UNIPARC:UPI000017761F; GB:M69181; NID:g641957
R;Weir, L.
submitted to the EMBL Data Library, August 1995
A;Reference number: H00753
A;Accession: G02055
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-81 <WEI>
A;Cross-references: UNIPARC:UPI000006EC6B; EMBL:U34304; NID:g1143217; PIDN:AAA84880.1; P
C;Genetics:
A;Gene: GDB:MYH10
A;Cross-references: GDB:127350; GDB:G00-127-350; OMIM:160776
A;Map position: 17p13-17p13
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F;88-771/Domain: myosin motor domain homology <MMO>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;559-572/Region: actin binding #status predicted
F;633-647/Region: actin binding #status predicted
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;701,711/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:              9.52e-10        Length:         1976
Score:                  314.00          Matches:        146
Percent Similarity:     40.64%          Conservative:   108
Best Local Similarity:  23.36%          Mismatches:     179
Query Match:            12.63%          Indels:         192
DB:                     2               Gaps:           26

US-09-837-344-41 (1-1482) x A59252 (1-1976)
```

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||   |||   |||   |||   |||:::   ::::::   |||   ::::::|||:::
Db       1122 GluLeuGlnGluAspPheGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLys 1141

Qy         64 AGCGATTTAGAACAAGATAGACTTGCT---AAAGAAAAGTTACAAGAG------------ 108
                ||||||   ::::::      |||   |||   :::|||::::::
Db       1142 ArgAspLeuSerGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThr 1161

Qy         109 ------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 162
                   |||   :::|||      :::|||      ::::::::::::   |||
Db       1162 ThrAlaAlaGlnGlnGluLeuArgThrLysArg------GluGlnGluValAlaGluLeu 1179

Qy         163 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 198
               :::      ||||||:::|||   :::      |||
Db       1180 LysLysAlaLeuGluGluGluThrLysAsnHisGluAlaGlnIleGlnAspMetArgGln 1199

Qy         199 ------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 246
                         |||:::|||            |||||||||   :::|||   |||
Db       1200 ArgHisAlaThrAlaLeuGluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLys 1219

Qy         247 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT------AGACTTGCTAAAGAA 300
                  |||::::::   :::       |||||   |||         ||||||   |||
Db       1220 AlaAsnLeuGluLysAsnLysGlnGlyLeuGluThrAspAsnLysGluLeuAlaCysGlu 1239

Qy         301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA--------- 351
                        ||||||          |||   |||   :::   |||:::
Db       1240 ValLysValLeuGlnGlnValLysAlaGluSerGluHisLysArgLysLysLeuAspAla 1259

Qy         352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
               ::::::||||||       :::   :::   :::       |||   :::   :::|||
Db       1260 GlnValGlnGluLeuHisAlaLysValSerGluGlyAspArgLeuArgValGluLeuAla 1279

Qy         412 GAACAACAAAGCGATTTAGAACAAGAG------------------------------- 438
                  |||:::   |||   |||:::   |||
Db       1280 GluLysAlaSerLysLeuGlnAsnGluLeuAspAsnValSerThrLeuLeuGluGluAla 1299

Qy         439 ---------------AGACTTGCTAAAGAA-----------------AAGTTGCAAGAA 465
                            :::   ||||||:::                 :::||||||:::
Db       1300 GluLysLysGlyIleLysPheAlaLysAspAlaAlaSerLeuGluSerGlnLeuGlnAsp 1319

Qy         466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
                  |||      |||::::::|||   |||   |||       |||      :::
Db       1320 ThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnLeuSerSerArgIleArg 1339

Qy         526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                  ||||||:::|||          |||       ||||||||||||||        |||
Db       1340 GlnLeuGluGluGlu---------LysAsnSerLeuGlnGluGlnGln---------Glu 1353

Qy         586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
               :::|||      |||::::::   |||::::::|||          |||:::   :::
Db       1354 GluGluGluGluAlaArgLysAsnLeuGluLysGlnValLeuAlaLeuGlnSerGlnLeu 1373

Qy         646 GCTGATACGAAAAAAAATTTAGAA---------------------------------- 669
               ||||||||||||||   ::::::
Db       1374 AlaAspThrLysLysLysValAspAspAspLeuGlyThrIleGluSerLeuGluGluAla 1393

Qy         670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
               :::||||||         |||       ||||||   |||       |||||||||
Db       1394 LysLysLysLeuLeuLysAsp------AlaGluAlaLeuSerGlnArgLeuGlu------ 1409

Qy         730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                                :::   :::|||
Db       1410 ----------------------------------------------GluLysAlaLeu 1413

Qy         790 CCTCAGGAC------AACAGAGGGAATAGTAGAGATTCCAAGGAAATATCT------ATA 837
                  |||         :::      :::|||      :::|||:::
Db       1414 AlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuThr 1433

Qy         838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA------------ 885
               ::::::      ::::::      :::   :::|||:::|||   :::
Db       1434 ValAspLeuAspHisGlnArgGlnValAlaSerAsnLeuGluLysLysGlnLysLysPhe 1453
```

```
Qy        886 ----------------------AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                                    :::   |||         :::   ||||||:::
Db       1454 AspGlnLeuLeuAlaGluGluLysSerIleSerAlaArgTyrAlaGluGluArgAspArg 1473

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA--------------- 969
                  :::      :::||||||   |||:::      :::
Db       1474 AlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGlu 1493

Qy        970 --------------------------------CAAAATCATACATTAGAGACAGTAAAT 996
                                               ||||||      |||
Db       1494 GluAlaLeuGluAlaLysGluGluPheGluArgGlnAsnLysGlnLeuArg--------- 1510

Qy        997 ATTTCTGATGTTAATGATTTT------------------------------------- 1017
                  :::|||:::   |||
Db       1511 ---AlaAspMetGluAspLeuMetSerSerLysAspAspValGlyLysAsnValHisGlu 1529

Qy       1018 ------------------------------------------------CAAATAAGT 1026
                                                              |||:::
Db       1530 LeuGluLysSerLysArgAlaLeuGluGlnGlnValGluGluMetArgThrGlnLeuGlu 1549

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              :::   |||||||||:::   |||      |||      |||      |||
Db       1550 GluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMet 1569

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                        :::|||:::      :::|||      :::   |||:::||||||
Db       1570 GlnAlaMetLysAlaGlnPheGluArgAspLeuGlnThrArgAspGluGlnAsnGluGlu 1589

Qy       1147 AAC------------ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
                           |||         :::|||||||||      :::|||
Db       1590 LysLysArgLeuLeuIleLysGlnValArgGluLeuGluAlaGluLeuGluAspGluArg 1609

Qy       1195 ----------------------------------------------AATTTAGATGATTTA 1209
                                                            :::|||   ||||||
Db       1610 LysGlnArgAlaLeuAlaValAlaSerLysLysLysMetGluIleAspLeuLysAspLeu 1629

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
              :::      ||||||   ::::::   :::      :::|||   ||||||:::   :::
Db       1630 GluAlaGlnIleGlu---AlaAlaAsnLysAlaArgAspGluValIleLysGlnLeuArg 1648

Qy       1270 AAATATGAA---AAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
              |||   :::   :::   ||||||                                |||
Db       1649 LysLeuGlnAlaGlnMetLysAsp----------------------------TyrGln 1658

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAG-----------GTTAATAAGGAAAAG 1374
                   :::::::::   :::      :::                  :::||||||
Db       1659 ArgGluLeuGluGluAlaArgAlaSerArgAspGluIlePheAlaGlnSerLysGluSer 1678

Qy       1375 GAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATC 1434
              ||||||   :::|||||||||                :::   ||||||||||||:::
Db       1679 GluLysLysLeuLysSerLeu-----------------GluAlaGluIleLeuGlnLeu 1692

Qy       1435 GTGGATGAGTTATCT 1449
                 :::||||||:::
Db       1693 GlnGluGluLeuAla 1697

RESULT 21
T14867
interaptin - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T14867
R;Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A;Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dictyo
ts.
A;Reference number: Z18248; MUID:98365468; PMID:9700162
A;Accession: T14867
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1738 <RIV>
A;Cross-references: UNIPROT:O76329; UNIPARC:UPI000007F02C; EMBL:AF057019; NID:g3549260;
C;Genetics:
A;Gene: abpD
A;Introns: 173/2; 1680/1

Alignment Scores:
Pred. No.:              1.03e-09        Length:         1738
Score:                  313.50          Matches:        127
Percent Similarity:     46.86%          Conservative:   127
Best Local Similarity:  23.43%          Mismatches:     171
Query Match:            12.61%          Indels:         117
DB:                     2               Gaps:           24

US-09-837-344-41 (1-1482) x T14867 (1-1738)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA--------------GAA 45
              |||:::   ||||||   |||   ::::::|||:::   :::                |||
Db       1012 GlnGlnLeuGlnSerGlnLeuAsnGluGlnArgGlnGlnGlnSerAsnGlnLeuSerGlu 1031

Qy         46 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
              |||   |||:::             :::|||:::::::::      :::::::::   |||
Db       1032 LysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGlnPheAspGlnLysGluGlnGln 1051

Qy        106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA---AAGTTGCAAGAACAA 162
                 :::|||      :::|||   :::          ||||||   ::::::|||:::
Db       1052 LeuLysGlnGlnSerIleGluAsnAspLeuPheGluLysGluAsnGlnIleGlnGlnLeu 1071

Qy        163 CAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 219
              ||||||   |||   ||||||   ::::::          :::|||   |||:::
Db       1072 GlnSerGlnLeuAsnGluGlnArgGlnGlnGlnSerAsnGlnLeuSerGluLysAspGln 1091

Qy        220 GATTTAGAACAA--------------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 264
              |||   |||                |||        ||||||:::      :::|||
Db       1092 GlnLeuAsnGlnLeuIleGluLysAsnGluSerAspGlnLysGluGlnGlnLeuLysGln 1111

Qy        265 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA---AAGTTACAAGAGCAGCAAAGC 321
              ||||||   :::|||   |||   :::   ||||||   ::::::|||:::   |||
Db       1112 GlnSer---IleGluAsnAspLeuIleGluLysGluAsnGlnIleGlnGlnLeuGlnLeu 1130

Qy        322 GATTTAGAACAAGAGAGACGT---------------------------------------- 342
              |||   ::::::|||:::
Db       1131 GlnLeuAsnGluGlnArgGlnLeuGlnSerGluValSerIleAspAsnAspLysIleLeu 1150

Qy        343 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA------------GAACAAGAG 387
                 :::::::::|||::::::   ||||||||||||          :::::::::
Db       1151 GluLeuGluLysGlnLeuLysGlnCysGlnSerAspLeuLeuLysLeuAsnAspGluLys 1170

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
              ::::::      ::::::||||||::::::|||   :::   :::|||   :::|||
Db       1171 GlnGlnGlnAspLysGlnLeuGlnAspLysGlnIleGluPheAspGlnLeuGlnLeuThr 1190

Qy        448 ------------------------------------------------------------AAA 450
                                                                          |||
Db       1191 PheAsnGlnPheLysAsnAspLysAspSerGlnPheIleGlnLeuGlnAspAspGlnLys 1210

Qy        451 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
              ::::::||||||      |||   ||||||   |||           |||::::::
Db       1211 GlnGlnLeuGlnSerIleGlnGlnAspLeuAsnGln---------LeuLysGlnGluAsn 1227

Qy        511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
              ||||||::::::      |||   ::::::            :::   :::|||   :::
Db       1228 GlnGluLysGluLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnPheGlu 1247

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAGCAGCAAAGA 627
                 :::   |||::::::          |||   ||||||||||||      |||:::
Db       1248 AsnGlnGluLysGluLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGln 1267

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAA---------AAAAATTTAGAAAGAAAAAAG 678
              :::|||   |||         :::   :::         |||         |||
Db       1268 AsnLeuAsnGlnLeuAsnAspGluAsnGlnGluLysValLysGlnPheSerGluLysAsp 1287

Qy        679 GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAA 738
              |||           :::   :::||||||                          :::
Db       1288 GluLysLeuGlnSerIleGlnGlnAspLeuAsn----------------------Gln 1299
```

```
Qy        739 CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGAC 798
              |||      ||||||:::                                  ::::::
Db       1300 LeuLysGlnGluAsnGlnGlu--------------------------------LysGlu 1308

Qy        799 AACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA---------ATAGAAAAAACA 849
                 :::   :::   :::|||   |||      ||||||         :::   :::
Db       1309 LysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAspLeuAsnGlnLeu 1328

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              |||   :::   |||      |||   :::                            |||
Db       1329 AsnAspAspGlnIleLysLysAsnGluLys--------------------------Leu 1339

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              :::||||||:::::      :::|||   :::|||      :::
Db       1340 LysGluLysGluGluGlnLeuLeuLysLeuGlnGlnAspPheAsnAspGlnGlnSerGln 1359

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
              |||         ||||||                  |||      |||      |||
Db       1360 GlnLeuLysGlnLeuGluGluLysLeuSerGluLysGluAsnGlnLeuGlnGlnLeuLys 1379

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                 |||:::||||||                ::   |||      :::::::::   :::
Db       1380 GlnGluAsnGluIle---------------AsnGlnLeuAsnGlnGlnGlnGlnSerAsn 1394

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
              |||   :::   :::   |||      :::                 |||:::::::::
Db       1395 GluIleIleGlnGlnLeuLysAspGlnLeuLeuLysGlnGlnGlnGlnGluGlnGlnGlu 1414

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
                    ||||||:::|||   ||||||::::::   |||   |||
Db       1415 AsnAsnAsnGluLysGluIleGluArgLeuIleGlnGluIleGluGlnLeuLysGlnGln 1434

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA--------- 1260
                 |||   ||:::::::|||      |||||||||   :::::::||||||
Db       1435 GlnGlu---IleAspGlnSer------GluLeuSerAsnLysGluIleLysIleGlnThr 1451

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
                    :::::::::::::::          |||      |||
Db       1452 ThrGlnGlnGluPheAspGlnLeuSerHisAsnArgSerLys------------------ 1465

Qy       1321 TATGATGAG---CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
                 |||:::   |||:::::::::   :::      :::::::::::::::   |||:::
Db       1466 ---AspGlnLeuHisLeuGlnGlnLeuGln------GlnGluLeuAspGlnLeuLysGln 1482

Qy       1378 AAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTG 1437
                 |||                        |||   |||::::::      :::::::::
Db       1483 SerPhe--------------------AspAspGlnAspHisGlnPheLysLysValIle 1495

Qy       1438 GATGAG 1443
              ||||||
Db       1496 AspGlu 1497

RESULT 22
A33977
myosin heavy chain, nonmuscle - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A33977; S06116; A43422
R;Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A;Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A;Reference number: A33977; MUID:90046668; PMID:2813355
A;Accession: A33977
A;Molecule type: mRNA
A;Residues: 1-1959 <SHO>
A;Cross-references: UNIPROT:P14105; UNIPARC:UPI000012FB7F; GB:M26510; NID:g212382; PIDN:
R;Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A;Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A;Reference number: S06116; MUID:90032648; PMID:2806244
A;Accession: S06116
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 716-1008 <KAT>
A;Cross-references: UNIPARC:UPI0000173E31; GB:X17589
A;Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114
R;Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A;Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate
A;Reference number: A43422; MUID:92381096; PMID:1512291
A;Accession: A43422
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1900-1959 <HOD>
A;Cross-references: UNIPARC:UPI000008A46C
A;Experimental source: brush border
A;Note: sequence extracted from NCBI backbone (NCBIP:111947)
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleoti
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1936/Domain: coiled coil #status predicted <COI>
F;837-1277/Region: S2
F;1278-1959/Region: light meromyosin
F;1937-1959/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694,704/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                1.14e-09      Length:          1959
Score:                    312.50        Matches:         140
Percent Similarity:       41.75%        Conservative:    108
Best Local Similarity:    23.57%        Mismatches:      157
Query Match:              12.57%        Indels:          189
DB:                       1             Gaps:            26

US-09-837-344-41 (1-1482) x A33977 (1-1959)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||:::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAA--------------------AGCGATTTAGAACAAGAT----------------- 81
              :::                          ||||||::|||
Db        966 LysValThrThrGluAlaLysLeuLysLysLeuGluGluAspValIleValLeuGluAsp 985

Qy         82 ---------AGACTTGCTAAAGAAAAG------TTACAAGAGCAGCAAAGCGATTTA--- 123
                    :::|||||||||||||||      |||:::::::::   |||:::
Db        986 GlnAsnLeuLysLeuAlaLysGluLysLysLeuGluAspArgMetSerGluPheThr 1005

Qy        124 --------------GAACAAGAGAGA---------CTTGCTAAAGAAAAGTTGCAAGAA 159
                            |||:::|||:::         |||||||||   |||   :::
Db       1006 ThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeuLysAsnLysHis 1025

Qy        160 CAACAA------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 213
              :::         :::|||||||||:::       |||   :::||||||         |||
Db       1026 GluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys---------Gln 1042

Qy        214 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273
              :::   :::||||||:::   ||||||         ||||||:::          ||||||
Db       1043 ArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGlyAspSerSerAsp 1059

Qy        274 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA--- 330
              |||      |||::::::|||      :::||||||   |||   ::::::|||:::
Db       1060 Leu---HisAspGlnIleAla------GluLeuGlnAlaGlnIleAlaGluLeuLysIle 1076

Qy        331 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--- 387
              |||      :::   :::|||:::||||||         :::   :::|||:::|||
Db       1077 GlnLeuSerLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgValGluGluGluAla 1096

Qy        388 ----------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
```

```
Db       1097 AlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGlnIleThrGluLeu 1116

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477

Db       1117 GlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAsp 1136

Qy        478 TTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG---------------------- 510

Db       1137 LeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAla 1156

Qy        511 ------CAAGAACAACAAAGCGATTTAGAACAAGAG---------------------- 540

Db       1157 AlaGlnGlnGluLeuArgSerLysArgGluGlnGluValThrValLeuLysLysThrLeu 1176

Qy        541 ---------------------------------------------AGACGTGCTAAAGAA 555

Db       1177 GluAspGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGln 1196

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 615

Db       1197 AlaIleGluGluLeuAlaGluGlnLeuGluGlnThrLysArgValLysAlaAsnLeuGlu 1216

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---------------------- 654

Db       1217 LysAlaLysGlnAlaLeuGluSerGluArgAlaGluLeuSerAsnGluValLysValLeu 1236

Qy        655 ---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGAT 705

Db       1237 LeuGlnGlyLysGlyAspAlaGluHisLysArgLysLysValAspAlaGlnLeuGlnGlu 1256

Qy        706 TTATATGGTCGTTTA---------------------------------GAAATACCA 729

Db       1257 LeuGlnValLysPheThrGluGlyGluArgValLysThrGluLeuAlaGluArgValAsn 1276

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789

Db       1277 LysLeuGlnValGluLeuAspAsnValThrGlyLeuLeu--------------------- 1289

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849

Db       1290 ------------------AsnGlnSerAspSerLysSerIleLysLeu------------ 1299

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909

Db       1299 ------------------------------------------------------------ 1299

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969

Db       1300 ------AlaLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeu 1317

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029

Db       1318 GlnGluGluThrArgLeuLysLeuSerPheSerThr------------LysLeuLysGln 1333

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089

Db       1334 ThrGluAspGluLysAsnAla---LeuLysGluGlnLeuGluGluGluGluGluAlaLys 1352

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149

Db       1353 ArgAsnLeuGluLysGlnIleSerValLeuGln-----------------GlnGlnAla 1366

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209

Db       1367 ValGluAlaArgLysLysMetAspAspGlyLeuGlyCysLeuGluIleAlaGluGluAla 1386

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAAAAAATAAAA 1260

Db       1387 LysLysLysLeuGlnLysAspLeuGluSerLeuThrGlnArgTyrGluGluLysIleAla 1406

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------------- 1290

Db       1407 AlaTyrAspLysGluLeuGluLysThrLysThrArgLeuGlnGlnGluLeuAspAspIleAla 1426

Qy       1291 ----------------------------AATAATTTTAAACCAAATGATAAAAGTTTG 1320

Db       1427 ValAspLeuAspHisGlnArgGlnThrValSerAsnLeuGluLysLysGlnLysLys--- 1445

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAAT-------------------------GAT 1353

Db       1446 PheAspGlnLeuLeuAlaGluGluLysAsnIleSerAlaLysTyrAlaGluGluArgAsp 1465

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTG 1395

Db       1466 ArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu 1479
```

RESULT 23
B43402
myosin heavy chain-B, neuronal - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Dec-2004
C;Accession: B43402; A43402
R;Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A;Reference number: A43402; MUID:92388144; PMID:1355479
A;Accession: B43402
A;Molecule type: mRNA
A;Residues: 1-2007 <TAK>
A;Cross-references: UNIPROT:Q02015; UNIPARC:UPI00000FC6BE; GB:M93676; NID:g212448; PIDN:
A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A;Accession: A43402
A;Molecule type: mRNA
A;Residues: 1-211;222-631;653-2007 <TA2>
A;Cross-references: UNIPARC:UPI0000173E32; UNIPARC:UPI0000173E33; UNIPARC:UPI0000173E34;
A;Note: sequence extracted from NCBI backbone (NCBIN:112864)
C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F;1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS
F;88-802/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;212-221/Region: alternatively spliced segment 1 #status experimental
F;559-593/Region: actin binding #status predicted
F;632-652/Region: alternatively spliced segment 2 #status experimental
F;692-714/Region: actin binding #status predicted
F;875-2007/Domain: coiled coil #status predicted <COI>
F;875-1315/Region: S2
F;1316-2007/Region: light meromyosin
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;732,742/Active site: Cys #status predicted
F;1954/Binding site: phosphate (Thr) (covalent) #status predicted
F;1987/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:

Pred. No.:	1.21e-09	Length:	2007
Score:	312.00	Matches:	136
Percent Similarity:	44.68%	Conservative:	137
Best Local Similarity:	22.26%	Mismatches:	164
Query Match:	12.55%	Indels:	174
DB:	1	Gaps:	28

US-09-837-344-41 (1-1482) x B43402 (1-2007)

```
Qy          10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69

Db         903 GluAlaGluLeuGluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIle 922

Qy          70 TTAGAACAAGAT---------------------------------------------- 81

Db         923 LeuAlaGluGlnLeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAla 942
```

```
Qy        82 AGACTTGCT------AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA------------ 123
             |||||||||      |||:::::::|||:::|||       ||||||
Db       943 ArgLeuAlaAlaLysLysGlnGluLeuGluGluIleLeuHisAspLeuGluSerArgVal 962

Qy       124 ------GAACAAGAGAGA---------CTTGCTAAAGAAAAG------------------ 150
                 |||:::||||||         |||      ||||||
Db       963 GluGluGluGluGluArgAsnGlnIleLeuGlnAsnGluLysLysLysMetGlnGlyHis 982

Qy       151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
             :::|||:::   :::      |||:::::|||      |||:::::|||||||||
Db       983 IleGlnAspLeuGluGluGlnLeuAspGluGluGluGlyAlaArgGlnLysLeuGlnLeu 1002

Qy       211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG-----------TTGCAA 258
             ::::::          |||   :::   :::      |||:::           |||:::
Db      1003 GluLysValThrAlaGluAlaLysIleLysLysMetGluGluGluIleLeuLeuLeuGlu 1022

Qy       259 GAACAACAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
             :::|||   |||      |||:::::   :::|||   ::::::::::::   |||
Db      1023 AspGlnAsnSerLysPheLeuLysGluLysLysLeuMetGluAspArgIleAlaGluCys 1042

Qy       316 CAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA---------------------- 348
                |||   |||   |||:::|||   :::||||||
Db      1043 ThrSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAlaLysLeuLysAsnLys 1062

Qy       349 --------------------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
                                       |||:::|||:::::::::::      :::
Db      1063 GlnGluMetMetIleThrAspLeuGluGluArgLeuLysLysGluGluLysThrArgGln 1082

Qy       382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
             :::      :::||||||   ||||||      :::   :::||||||   |||::::::
Db      1083 GluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAspLeu---GlnAspGln 1101

Qy       442 CTTGCTAAA--------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 486
             :::|||:::              |||:::|||:::   |||   :::      |||:::
Db      1102 IleAlaGluLeuGlnAlaGlnIleGluGluLeuLysIleGlnLeuAlaLysLysGluGlu 1121

Qy       487 GAG--------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531
             |||              |||     |||:::      ::::::   :::   |||
Db      1122 GluLeuGlnAlaAlaLeuAlaArgGlyAspGluGluAlaValGlnLysAsnAsnAlaLeu 1141

Qy       532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591
             :::      |||      :::   :::::   |||   |||   |||||||||   |||
Db      1142 LysValIleArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspLeuGluSerGlu 1161

Qy       592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTA------------------ 633
             :::   ::::::   |||   ::::::|||:::|||||||||
Db      1162 LysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAla 1181

Qy       633 ------------------------------------------------------------ 633

Db      1182 LeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThr 1201

Qy       634 ------GAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
                   ||||||      |||:::   ||||||   :::|||   :::   ::::::
Db      1202 LysArgGluGlnGluValAlaGluLeuLysLysAlaIleGluGluGluThrLysAsnHis 1221

Qy       688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
             :::          :::::::::      |||                  ||||||
Db      1222 GluAlaGlnIleGlnGluIleArgGlnArgHisAlaThrAlaLeuGluGluLeuSerGlu 1241

Qy       748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
             :::   |||:::                     :::::::::|||   ::::::   :::|||
Db      1242 GlnLeuGluGlnAlaLysArgPhe------LysAlaAsnLeuGluLysAsnLysGlnGly 1259

Qy       808 AATAGTAGAGATTCCAAG------------GAAATATCTATAATAGAAAAAACAAATAGA 855
                    |||:::|||             |||:::   ::::::::::::
Db      1260 LeuGluSerAspAsnLysGluLeuAlaCysGluValLysValLeuGlnGlnValLysAla 1279

Qy       856 GAATCT--------------------------------------ATTACAACAAATGTT 876
             ||||||                                      :::|||      |||
Db      1280 GluSerGluHisLysArgLysLysLeuAspAlaGlnValGlnGluLeuThrAlaLysVal 1299
```

```
Qy       877 ---GAAGGACGAAGG---------------------------------------------- 888
                 ||||||   |||
Db      1300 ThrGluGlyGluArgLeuArgValGluLeuAlaGluLysAlaAsnLysLeuGlnAsnGlu 1319

Qy       889 --------------------------GATATACATAAAGGACATCTTGAAGAAAAGAAA 921
                                        :::      |||      ::::::      |||
Db      1320 LeuAspAsnValSerSerLeuLeuGluGluAlaGluLysLysGlyIleLysPheAlaLys 1339

Qy       922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
             |||   :::            |||   :::|||             :::|||      |||
Db      1340 AspAlaAlaSerLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThr 1359

Qy       982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
                 :::   :::|||:::|||                 :::|||   :::   |||:::|||
Db      1360 ArgGlnLysLeuAsnLeuSerSer------------ArgIleArgGlnLeuGluGluGlu 1375

Qy      1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
                                 ::::::|||   :::::::::|||:::::::|||:::
Db      1376 --------------LysAsnAsnLeuGlnGluGlnGlnGluGluGluGluGluAlaArg 1390

Qy      1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
             :::                              |||      :::::::::   :::   :::
Db      1391 Lys-----------------------AsnLeu---GluLysGlnMetLeuAlaLeuGln 1401

Qy      1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT------GAA 1215
                 :::|||   :::      :::|||   :::      |||||||||          |||
Db      1402 AlaGlnLeuAlaGluAlaLysLysLysValAsp------AspAspLeuGlyThrIleGlu 1419

Qy      1216 GGAATAGAA--------------------AAATCATCAGAAGAATTATCT--------- 1245
             |||:::|||                   |||      |||   ||||||
Db      1420 GlyLeuGluGluAsnLysLysLysLeuLeuLysAspMetGluSerLeuSerGlnArgLeu 1439

Qy      1246 GAAGAAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
             |||||||||                  |||   |||||||||||:::         :::
Db      1440 GluGluLysAlaMetAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGlu 1459

Qy      1306 AATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTT 1362
                |||      :::   |||   :::|||   ::::::      ::::::      :::
Db      1460 LeuAspAspLeuMetValAspLeuAspHisGlnArgGlnIleValSerAsn------Leu 1477

Qy      1363 AATAAGGAAAAGGAAAAATTCATAAAATCATTG 1395
                |||:::::::::||||||   :::   |||
Db      1478 GluLysLysGlnLysLysPheAspGlnMetLeu 1488
```

RESULT 24
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: UNIPROT:Q07569; UNIPARC:UPI00000802D3; EMBL:L03534; NID:g1850912; PI
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Alignment Scores:

Pred. No.:	1.27e-09	Length:	2139
Score:	311.50	Matches:	128
Percent Similarity:	46.20%	Conservative:	127
Best Local Similarity:	23.19%	Mismatches:	184
Query Match:	12.53%	Indels:	113
DB:	2	Gaps:	25

US-09-837-344-41 (1-1482) x T18296 (1-2139)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT------------------------ 36
Db       1014 GluGluGluLysAspAspLeuGluGlnAspArgAlaAspValSerAlaThrLysAspAsp 1033

Qy         37 ----------------------------------------GCTAAAGAAAAGTTGCAA 54
Db       1034 IleAlaLysLysLeuAsnLysIleThrIleGluCysGluAspAlaLysAspGluIleAla 1053

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGA---------------CTTGCTAAAGAA--- 96
Db       1054 LysLeuGluGlnGluLeuGluAspGluGluAsnLysAsnLysAspLeuThrAsnGluLeu 1073

Qy         97 ---------------AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCT 141
Db       1074 GlnGlnThrGlnLeuLysLeuGlyGluThrGluLysSerLeuAlaAlaGlnValAlaAla 1093

Qy        142 AAAGAAAAGTTGCAAGAACAACAA------------AGCGATCTAGAACAAGAGAGACGT 189
Db       1094 ThrLysLysAlaSerAspGluArgAspThrLeuSerGlnAsnLeuGluAsnGluLysLeu 1113

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG------------AGA 237
Db       1114 ThrThrLysAsnLeuThrLysThrLysAlaAspLeuGluLysLysIleSerGlyLeuLys 1133

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCT 294
Db       1134 GlnAspTyrGluAspLeuGluAspAspLysAsnLysIleGluGlyAspLeuArgAsnAla 1153

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 354
Db       1154 GlnArgLysIleLysGluLeuAspAspGluIleThrLysGlyAlaAspValSerGlnTyr 1173

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
Db       1174 LeuGlnLysGlnLysGluGluTyrGluSerGln---------IleAlaLysMetGlnGlu 1190

Qy        415 CAACAAAGCGATTTA---------------GAACAAGAGAGACTTGCTAAAGAAAAG--- 456
Db       1191 GluLysGluAlaIleGlyAsnAspValLysAsnLysGluLysThrIleLysGluLysGlu 1210

Qy        457 ------TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
Db       1211 LeuGluIleGlnSerLeuGlnGluLysLeuAspGluThrGluValGluLysGluAspAla 1230

Qy        511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
Db       1231 GluLysLysLysLysGluIleGluLysGluMetLysAla---------LeuGlnGluGlu 1247

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
Db       1248 LysGluAsnValGluSerSerLysAsnSerThrGluLys---------AspLysLysLys 1264

Qy        631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
Db       1265 LeuGluAspAsnLeuLysAspThrGlnLysLysLeuAspAspMetThrAlaAspAsnGlu 1284

Qy        691 ATATTA---------GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
Db       1285 LysLeuLysAlaLysAlaLysAspLeuGluAlaGlnLeuAsn-----------GluVal 1300

Qy        742 CCATCAGAAAATGAACGTGGATAT---------TATATACCACATCAATCTTCTTTACCT 792
Db       1301 GlnAspAsnHisGluLysAlaValAlaAspAlaGluLeuLeuAsnLysLysLysAlaGln 1320

Qy        793 CAGGACAACAGAGGGAATAGTAGAGAT-----------------TCCAAGGAAATATCT 834
Db       1321 SerAspLysGluLeuAsnSerLeuLysAlaGluLeuGluAlaLeuThrLysAlaLysSer 1340

Qy        835 ATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATA 894
Db       1341 ValValGluSerLysAsnLysAspSer-------------------------GluAsn 1351

Qy        895 CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGAT 954

Db       1352 GluLysAlaAlaLeuSerGluGluIleAspGlnAla------AsnGluLysLeuLysAsn 1369

Qy        955 AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
Db       1370 IleGlnAlaAspLeuArgLys---------AlaThrAlaAspLeuGlnGluAlaAsnGlu 1386

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGAT 1074
Db       1387 --------------LysLysAlaGluValGluAlaGlnArgAspLysLeuValAlaAsp 1401

Qy       1075 GAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTC 1134
Db       1402 AsnLysLysMetThrLysThrLeuGluGluIleLysAlaArgAspGluGluAsnThrTyr 1421

Qy       1135 CAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA---GATTTGATAGAGAAAAAT 1191
Db       1422 LysValGluAsnTyrGluLysValLeuLysArgLysGluAlaAspLeuGluGluAlaAsn 1441

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
Db       1442 GluAsnLeuAsp---------------IleGluLysLysAspArgMetAsnLysGluLys 1456

Qy       1252 AAAATAAAAAAAGGAAAG---AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
Db       1457 GlnValLysLysLeuGluGlyGluLeuLysGluThrLysAsp---------LysLeuAsn 1473

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
Db       1474 AlaAlaIleAlaGluLysAspSerIle------PheThrAlaLysLysGlnSerAspAla 1491

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
Db       1492 AspLeuGluGluLeuAsnLysThrVal--------------------GluGluHisAsp 1504

Qy       1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
Db       1505 GluValValAlaLysLeuAsnThrGlnIleThrLys 1516
```

RESULT 25
B71609
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: B71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-665 <GAR>
A;Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; GB:AE001362; NID
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0680w

Alignment Scores:

Pred. No.:	1.74e-09	Length:	665
Score:	310.50	Matches:	115
Percent Similarity:	47.02%	Conservative:	122
Best Local Similarity:	22.82%	Mismatches:	172
Query Match:	12.49%	Indels:	95
DB:	2	Gaps:	25

US-09-837-344-41 (1-1482) x B71609 (1-665)

```
Qy          1 CAAGAACAACAAAGCGAT---------CTAGAACAAGAGAGACGTGCTAAAGAA------ 45
Db        210 GluAsnLysLysSerAspAspHisLysValGluGluAsnLysLysSerAspAspHisLys 229
```

```
Qy        46 --------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 90
                           |||::::::|||   :::   |||   |||:::|||:::
Db       230 IleGluGluValLysLysValGluGluHisGluGluAspGluGluGluAspLys------ 247

Qy        91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 150
             |||||||||     |||   ::::::      |||                 |||::::::
Db       248 LysGluLysLysSerGluAsnLysAsnLysAspGluAsn------------LysAspGlu 263

Qy       151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
               |||         ::::::      |||             :::   ::::::|||
Db       264 AsnAspGluAspAsnAspGluIleSerAspGluAspGluValAspAspAspValGluGlu 283

Qy       211 ---------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTG 255
                  :::      |||::::::   :::::::::      :::|||          |||
Db       284 AspLysAsnGluAsnAspAspIleAspAspAspLysLysGluThrAspLysThrHisLeu 303

Qy       256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
             :::|||:::::::::::::   :::         :::   :::|||           :::
Db       304 GluGluGluGluAsnGluIleIleGluLysGluPheSerAspLysLysLysAsnGlyLys 323

Qy       316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
               |||   ::::::|||   :::::::|||:::      :::|||:::       |||
Db       324 AsnLysAspThrLysLysGlu---LysSerLysAspThrGluLysGluLysSerLysAsp 342

Qy       376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
             :::|||:::|||   ::::::|||:::|||   :::|||:::       |||   |||:::
Db       343 IleGluLysGlu---LysSerLysAspLysGluLysGluLysSerLysAspLysGluLys 361

Qy       436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
               |||:::      |||:::|||   :::|||:::      |||:::|||:::|||:::
Db       362 GluLys---GlyLysAspLysGluLysGluLysSerLysAspIleGluLysGluLys--- 379

Qy       496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA---------CGT 546
                                 |||:::      |||:::|||:::|||:::
Db       380 ------------------GluLysAspLysAspIleGluLysGluLysSerLysAspThr 393

Qy       547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAA 603
             ||||||||||||      |||:::   |||:::|||:::|||:::   :::
Db       394 AlaLysGluLys------GluLysAspLysAspIleGluLysGluLysSerLysAspMet 411

Qy       604 GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAT 663
             |||||||||:::   :::|||   |||   |||:::::::|||   |||   :::|||
Db       412 GluLysLeuLysAsnLysGlnAsnAsp---GluLysLysLysAspAspAsnGluLys--- 429

Qy       664 TTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA 723
                ::::|||   :::      ||||||
Db       430 ------LysLysAsnAspLysGlnAspIle---------------------------- 437

Qy       724 ATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCT 783

Db       437 ------------------------------------------------------------ 437

Qy       784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAA 843
                       |||:::      :::      |||   :::||||||          |||
Db       438 ---------HisAspAspAsnAspAspGluAsnAspMetGluGluIleGlu------Glu 452

Qy       844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903
                ::::   |||              ::::::|||   ::::::          ||||||
Db       453 AsnAspAspGluGluAspGluAspGluAspMetGluAsnLysLysLysLysLysLysGly 472

Qy       904 CATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
                           ::::::|||:::      :::      :::::::::
Db       473 Lys---AsnGlyAsnGluAsnGlyAsnGluAsnGlySerGluAsnGlyAsnGluAsnGly 491

Qy       964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
             :::      :::      |||   |||   |||:::   |||:::   :::
Db       492 AsnGluAsnGlyAsnGluAsnGluAsnLysAsnGluSerGlu---AsnGluAsnGlu--- 509

Qy      1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAA 1083
             ::::::   |||:::|||          |||   ::::::             |||   |||
Db       510 AsnGluAsnGluAsnGluAsnGlyAsnGluAsnGluAsnGlu------LysGluAsnGlu 527

Qy      1084 GATGATGAAGACTTAGACGAATTTAAGCCTATTGTG--------------CAATATGAC 1128
```

```
                |||:::::::::   |||   :::   :::                    |||:::
Db       528 LysAspLysAsnIleLysGluIleGluAsnValThrAsnAlaAsnLysGluAsnTyrGlu 547

Qy      1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                                  |||   |||   |||          |||:::
Db       548 LysIleAsnLysAsnSerGluIleThrIleThrLysSerAsnIleAspIleTyrAsnAsn 567

Qy      1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                |||   |||   |||               |||:::|||   :::      :::
Db       568 AsnArgAsnAsnAsp--------------IleAspLysValAsnAsnHisIlePheThr 582

Qy      1249 GAAAAAATAAAAAAAGGAAAGAAATAT---GAAAAAACAAAGGATAAT---------AAT 1296
                :::   ||||||          :::   |||:::   |||   |||          |||
Db       583 AsnGlnGlnLysLysHisAsnLeuHisAsnGluGlnAsnLysPheAsnGluThrLeuAsn 602

Qy      1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                   |||   |||:::   |||:::|||          |||          |||
Db       603 ValSerThrAsnHisLysAsnHisTyrGluGluLysLysLysTyrGluSerAsnMetPhe 622

Qy      1357 CAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGAC 1416
                |||:::|||          |||   :::   |||:::      |||
Db       623 AsnValAspLysArgMetHisLysAsnLeuThrSerMetAspThrIleLeuHisAsnLeu 642

Qy      1417 AATGAAATTTTA 1428
                |||:::   |||
Db       643 AsnAspLysLeu 646
```

RESULT 26
A54514
glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: GARP
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004
C;Accession: A54514
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Silva, A.; Anders, R.F.; Kemp, D.J.
Mol. Biochem. Parasitol. 31, 199-202, 1988
A;Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich pro
A;Reference number: A54514; MUID:89040048; PMID:2903445
A;Accession: A54514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-678 <TRI>
A;Cross-references: UNIPROT:P13816; UNIPARC:UPI000012B0E0; GB:J03998; NID:g160298; PIDN
C;Genetics:
A;Introns: 25/3
C;Superfamily: histone H1
C;Keywords: tandem repeat

Alignment Scores:

Pred. No.:	1.74e-09	Length:	678
Score:	310.50	Matches:	113
Percent Similarity:	47.35%	Conservative:	119
Best Local Similarity:	23.06%	Mismatches:	169
Query Match:	12.49%	Indels:	89
DB:	2	Gaps:	16

US-09-837-344-41 (1-1482) x A54514 (1-678)

```
Qy        16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA---CAACAAAGCGATTTA 72
                |||   ::::::|||      :::::::::|||   ::::::   ||||||   ::::::
Db       237 AspAsnAspLysGluAsnIleSerGluAspLysLysGluAspHisGlnGlnGluGluMet 256

Qy        73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
                :::            ||||||:::   |||:::::::::          |||||||||
Db       257 LeuLysThrLeuAspLysLysGluArgLysGlnLysGluLysGluMetLysGluGlnGlu 276

Qy       133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
             ::::::   |||:::|||   :::::::::            |||:::|||::::::
Db       277 LysIleGluLysLysLysLysLysGlnGlu-----------GluLysGluLysLysLys 292

Qy       193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
             :::||||||   :::::::::|||         ::::::||||||:::   |||||||||
Db       293 GlnGluLysGluArgLysLysGlnGlu------LysLysGluArgLysGlnLysGluLys 310
```

```
Qy       253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
Db       311 GluMetLysLysGlnLysLysIleGluLysGluArgLysLysLysGlu------------ 326

Qy       313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
Db       327 --------------GluLysGluLysLysLysLysHisAspLysGluAsnGluGlu 341

Qy       373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAG----------------------TTG 408
Db       342 ThrMetGlnGlnProAspGlnThrSerGluGluThrAsnAsnGluIleMetValProLeu 361

Qy       409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
Db       362 ProSerProLeuThrAspValThrThrProGluGluHisLysGluGlyGluHisLysGlu 381

Qy       469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
Db       382 GluGluHisLysGluGlyGluHisLysGluGlyGluHisLysGluGluGluHisLysGlu 401

Qy       529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       402 ---GluGluHisLysLysGluGluHisLysSerLysGluHisLysSerLysGlyLysLys 420

Qy       589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db       421 AspLysGlyLysLysAspLysGlyLysHisLysLysAlaLysLysGluLysValLysLys 440

Qy       649 GATACGAAAAAA----------------------------------------------AATTTA 666
Db       441 HisValValLysAsnValIleGluAspGluAspLysAspGlyValGluIleIleAsnLeu 460

Qy       667 GAAAGAAAAAAG------GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db       461 GluAspLysGluAlaCysGluGluGlnHisIleThrValGluSer------ArgProLeu 478

Qy       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db       479 SerGlnProGlnCysLysLeuIleAspGlu--------------------ProGluGln 491

Qy       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
Db       492 LeuThrLeuMetAspLysSerLys---------ValGluGluLysAsnLeuSerIleGln 508

Qy       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db       509 GluGlnLeuIleGlyThrIleGlyArgValAsnValValProArgArgAspAsnHisLys 528

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATA--------------------AAACCA 939
Db       529 LysLysMetAlaLysIleGluGluAlaGluLeuGlnLysGlnLysHisValAspLysGlu 548

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
Db       549 GluAspLysLysGluGluSerLysGluValGlnGluGluSerLysGluVal--------- 565

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
Db       566 ----------------------------GlnGluAspGluGluGluValGluGluAsp 575

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
Db       576 GluGluGluGluGluGluGluGluGluGluGluGluGluGlu--------------- 590

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
Db       591 --------------GluGluGluGlu--------------GluGluGluGluGluGlu 600

Qy      1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
Db       601 GluGluGluGluAspGluAspGluGluAspGluAspAspAlaGluGluAspGluAsp 620

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
Db       621 AspAlaGluGluAspGluAspAspAlaGluGluAspAspGluGluAspAspAspGlu 640

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
Db       641 GluAspAspAspGluAsp---GluAspGluAspGluGluAspGluGluGluGluGluGlu 659

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db       660 GluGluGluGluSerGluLysLysIleLys 669
```

RESULT 27
MWKW
myosin heavy chain B [similarity] - Caenorhabditis elegans
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Caenorhabditis elegans
C;Date: 13-Jun-1983 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: T20770; T21629; A93958; A93287; A21074; A02992
R;Kershaw, J.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19322
A;Accession: T20770
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1963 <WIL>
A;Cross-references: UNIPROT:O02244; UNIPARC:UPI0000061004; EMBL:Z81499; PIDN:CAB04089.1;
A;Experimental source: clone F11C3
A;Accession: T21629
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1963 <WI2>
A;Cross-references: UNIPARC:UPI0000061004; EMBL:Z83107; PIDN:CAB05505.1; GSPDB:GN00019;
A;Experimental source: clone F32A7
R;Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A;Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy ch
A;Reference number: A93958; MUID:83273600; PMID:6576334
A;Accession: A93958
A;Molecule type: DNA
A;Residues: 1-61,'EMSVIQ',65-376,'V',378-1963 <KAR>
A;Cross-references: UNIPARC:UPI000012FBDC; GB:J01050; NID:g156399; PIDN:AAA28124.1; PID:
R;McLachlan, A.D.; Karn, J.
Nature 299, 226-231, 1982
A;Title: Periodic charge distributions in the myosin rod amino acid sequence match cross
A;Reference number: A93287; MUID:82272395; PMID:7202124
A;Accession: A93287
A;Molecule type: DNA
A;Residues: 847-1333,'R',1335-1876,'L',1878-1963 <MCL>
A;Cross-references: UNIPARC:UPI0000173E26
R;Wills, N.; Gesteland, R.F.; Karn, J.; Barnett, L.; Bolten, S.; Waterston, R.H.
Cell 33, 575-583, 1983
A;Title: The genes sup-7 X and sup-5 III of Caenorhabditis elegans suppress amber nonsen
A;Reference number: A21074; MUID:83232892; PMID:6571695
A;Accession: A21074
A;Molecule type: DNA
A;Residues: 1873-1963 <WI3>
A;Cross-references: UNIPARC:UPI000016B912; GB:V01494; GB:J01049; NID:g6783; PIDN:CAA2473
C;Genetics:
A;Gene: unc-54; CESP:F11C3.3
A;Map position: 1
A;Introns: 21/3; 64/3; 111/3; 264/1; 525/3; 951/2; 1747/3; 1819/3; 1894/3
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co
F;84-775/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;662-684/Region: actin binding #status predicted
F;766-780/Region: actin binding #status predicted
F;848-1963/Domain: coiled coil #status predicted <COI>
F;848-1162/Region: S2
F;1163-1963/Region: light meromyosin
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted

F;702,712/Active site: Cys #status predicted

Alignment Scores:

Pred. No.:	1.54e-09	Length:	1963
Score:	310.00	Matches:	151
Percent Similarity:	37.22%	Conservative:	114
Best Local Similarity:	21.21%	Mismatches:	185
Query Match:	12.47%	Indels:	262
DB:	1	Gaps:	27

US-09-837-344-41 (1-1482) x MWKW (1-1963)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::|||      |||||||||          |||       |||         |||:::
Db        960 LysLysGlnIleGlnAspLeuGluMetSerLeuArgLysAlaGlu------SerGluLys 977

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||          |||   |||      ::::::::::::|||:::|||
Db        978 GlnSerLysAspHisGlnIleArgSerLeuGlnAspGluMetGlnGlnGlnAspGluAla 997

Qy        121 TTA-----------------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 156
              :::                      ||||||   :::   ::::::   ::::::
Db        998 IleAlaLysLeuAsnLysGluLysLysHisGlnGluGluIleAsnArgLysLeuMetGlu 1017

Qy        157 GAACAACAAAGCGATCTAGAACAA---------GAGAGACGTGCTAAAGAAAAGTTGCAA 207
              :::   ||||||:::   ::::::         :::   :::   |||   ||||||:::
Db       1018 AspLeuGlnSerGluGluAspLysGlyAsnHisGlnAsnLysValLysAlaLysLeuGlu 1037

Qy        208 GAACAACAAAGCGAT------------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
              :::          |||           ||||||:::|||:::||||||:::      |||
Db       1038 GlnThrLeuAspAspLeuGluAspSerLeuGluArgGluLysArgAlaArgAlaAspLeu 1057

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAG 312
                 :::|||:::      :::|||   :::   ::::::|||:::|||   :::   |||
Db       1058 AspLysGlnLysArgLysValGluGlyGluLeuLysIleAlaGlnGluAsnIleAspGlu 1077

Qy        313 ---------CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAA 360
                      |||:::   |||||||||          :::   ||||||   :::|||
Db       1078 SerGlyArgGlnArgHisAspLeuGluAsnAsnLeuLysLysLysGluSerGluLeuHis 1097

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
                 |||   ||||||   |||:::                |||||||||   |||
Db       1098 SerValSerSerArgLeuGluAspGluGlnAlaLeuValSerLysLeuGlnArgGlnIle 1117

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 |||                     :::   ::::::   |||   :::   :::|||
Db       1118 LysAsp-----------------GlyGlnSerArgIleSerGluLeuGluGluGluLeu 1131

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||   ||||||:::::::::   |||           :::|||||||||::::::|||
Db       1132 GluAsnGluArgGlnSerArgSerLysAlaAspArgAlaLysSerAspLeuGlnArgGlu 1151

Qy        541 ---AGACGTGCTAAAGAAAAGTTGCAAGAACAA------------CAAAGCGATTTAGAA 585
                  |||||||||   ||||||          :::   :::|||
Db       1152 LeuGluGluLeuGlyGluLysLeuAspGluGlnGlyGlyAlaThrAlaAlaGlnValGlu 1171

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA------ 639
                  ::::::   :::   :::|||   :::   :::|||||||||||||:::
Db       1172 ValAsnLysLysArgGluAlaGluLeuAlaLysLeuArgArgAspLeuGluGluAlaAsn 1191

Qy        640 ----------------------------AGGAAGGCTGATACG-------------- 654
                                         ||||||     |||
Db       1192 MetAsnHisGluAsnGlnLeuGlyGlyLeuArgLysLysHisThrAspAlaValAlaGlu 1211

Qy        655 --------------------------------AAAAAAAATTTAGAAAGAAAAAAGGAA 681
                                               |||      :::|||:::   |||
Db       1212 LeuThrAspGlnLeuAspGlnLeuAsnLysAlaLysAlaLysValGluLysAspLysAla 1231

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA-------------------- 720
                  ||||||||||||      :::|||
Db       1232 GlnAlaValArgAspAlaGluAspLeuAlaAlaGlnLeuAspGlnGluThrSerGlyLys 1251

Qy        721 ----------------------------GAAATACCAGCTATAGAACTTCCATCAGAA 750
                                                  |||:::         ||||||   |||:::
Db       1252 LeuAsnAsnGluLysLeuAlaLysGlnPheGluLeuGlnLeuThrGluLeuGlnSerLys 1271

Qy        751 ---AATGAACGTGGATATTATATACCACATCAATCTTCTTTA----------------- 789
                  :::|||:::          :::          :::||||||
Db       1272 AlaAspGluGlnSerArgGlnLeuGlnAspPheThrSerLeuLysGlyArgLeuHisSer 1291

Qy        790 ------------------------------------CCTCAGGACAACAGAGGGAAT 810
                                                      |||   |||:::
Db       1292 GluAsnGlyAspLeuValArgGlnLeuGluAspAlaGluSerGlnValAsnGlnLeuThr 1311

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGA--------------- 855
                    |||:::      |||   :::|||:::      |||
Db       1312 ArgLeuLysSerGlnLeuThrSerGlnLeuGluGluAlaArgArgThrAlaAspGluGlu 1331

Qy        855 ------------------------------------------------------------ 855

Db       1332 AlaArgGluArgGlnThrValAlaAlaGlnAlaLysAsnTyrGlnHisGluAlaGluGln 1351

Qy        856 ------GAATCTATTACAACAAATGTTGAAGGACGAAGG--------------------- 888
                    ||||||:::          :::||||||:::
Db       1352 LeuGlnGluSerLeuGluGluGluIleGluGlyLysAsnGluIleLeuArgGlnLeuSer 1371

Qy        889 ------------GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA 936
                          ||||||   :::                 :::   :::|||   :::|||
Db       1372 LysAlaAsnAlaAspIleGlnGlnTrpLysAlaArgPheGluGlyGluGlyLeuLeuLys 1391

Qy        937 CCAGAACAAAAAGAAGAT---------------------------------------- 954
                 ::::::   ||||||
Db       1392 AlaAspGluLeuGluAspAlaLysArgArgGlnAlaGlnLysIleAsnGluLeuGlnGlu 1411

Qy        955 ------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT---------ATT 999
                        :::|||:::   :::|||   :::||||||                  :::
Db       1412 AlaLeuAspAlaAlaAsnSerLysAsnAlaSerLeuGluLysThrLysSerArgLeuVal 1431

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGT-------------------------------- 1026
                 |||::::::|||   |||:::
Db       1432 GlyAspLeuAspAspAlaGlnValAspValGluArgAlaAsnGlyValAlaSerAlaLeu 1451

Qy       1027 ---------------------------------------------AAGTATGAGGATGAA 1041
                                                            |||      |||:::
Db       1452 GluLysLysGlnLysGlyPheAspLysIleIleAspGluTrpArgLysLysThrAspAsp 1471

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
              ::::::||||||   |||   :::   |||                  ||||||
Db       1472 LeuAlaAlaGluLeuAspGlyAlaGlnArgAspLeuArgAsnThrSerThrAspLeuPhe 1491

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA----------------- 1143
              :::   |||                   |||   |||:::|||
Db       1492 LysAlaLys-----------------AsnAlaGlnGluGluLeuAlaGluValValGlu 1505

Qy       1144 ------------GAAAACATAGGAATTTATAAAGAACTAGAAGATTTG------------ 1179
                         ||||||      :::   :::|||::::::||||||
Db       1506 GlyLeuArgArgGluAsnLysSerLeuSerGlnGluIleLysAspLeuThrAspGlnLeu 1525

Qy       1180 ------------------------------------------------------ATAGAG 1185
                                                                    ||||||
Db       1526 GlyGluGlyGlyArgSerValHisGluMetGlnLysIleIleArgArgLeuGluIleGlu 1545

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              |||   |||              |||||||||          :::::::::   |||
Db       1546 LysGluGluLeuGlnHisAlaLeuAspGluAla------GluAlaAlaLeuGluAlaGlu 1563

Qy       1246 GAAGAAAAAAATA------------------------------------------- 1257
              |||   |||:::
Db       1564 GluSerLysValLeuArgAlaGlnValGluValSerGlnIleArgSerGluIleGluLys 1583

Qy       1258 ---------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                      :::|||   :::::::::|||   |||:::   |||:::   :::
Db       1584 ArgIleGlnGluLysGluGluGluPheGluAsnThrArgLysAsnHisAlaArgAlaLeu 1603

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
              :::                                          |||   :::
```

```
Db       1604 Glu-----------------------------------------SerMetGlnAlaSerLeu 1610

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
Db       1611 GluThrGluAla-------------------------LysGlyLysAlaGluLeuLeu 1621

Qy       1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
Db       1622 ArgIleLysLysLysLeuGluGlyAspIleAsnGlu 1633

RESULT 28
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
A;Cross-references: UNIPROT:P08799; UNIPARC:UPI000012FBD3; GB:M14628; GB:M11938; NID:g16
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
A;Cross-references: UNIPARC:UPI0000177634
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
A;Cross-references: UNIPARC:UPI0000177635
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
F;1-818/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

Alignment Scores:
Pred. No.:                1.52e-09        Length:          2116
Score:                    310.00          Matches:         129
Percent Similarity:       44.83%          Conservative:    144
Best Local Similarity:    21.18%          Mismatches:      180
Query Match:              12.47%          Indels:          156
DB:                       2               Gaps:            26

US-09-837-344-41 (1-1482) x A26655 (1-2116)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        884 AspSerLysAspAlaLeuGluAlaGlnLysArgGluLeuGluIleArgValGluAspMet 903

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db        904 GluSerGluLeuAspGluLysLysLeuAlaLeuGluAsnLeuGlnAsnGlnLysArgSer 923

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db        924 ValGlu-----------------GluLysValArgAspLeuGluGluGluLeuGlnGlu 937

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Db        938 GluGlnLysLeuArgAsnThrLeuGluLysLeuLysLysLysTyrGluGluGluLeuGlu 957

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT-----------TTAGAA------CAA 282
Db        958 GluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGluLysIleLys 977

Qy        283 GATAGACTTGCTAAA------GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 336
Db        978 AspGluLeuGlnLysGluValGluGluLeuThrGluSerPheSerGluGluSerLysAsp 997

Qy        337 AGACGTGCTAAAGAAAAG-----------TTGCAAGAACAACAAAGCGAT--------- 375
Db        998 LysGlyValLeuGluLysThrArgValArgLeuGlnSerGluLeuAspAspLeuThrVal 1017

Qy        376 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
Db       1018 ArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGlnLysLysLysLeuGlu 1037

Qy        433 CAAGAG-----------------------------------------AGACTTGCTAAA 450
Db       1038 GluGluLeuLysGlnValGlnGluAlaLeuAlaAlaGluThrAlaAlaLysLeuAlaGln 1057

Qy        451 GAA-----------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------- 489
Db       1058 GluAlaAlaAsnLysLysLeuGlnGlyGluTyrThrGluLeuAsnGluLysPheAsnSer 1077

Qy        490 ----------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA------ 519
Db       1078 GluValThrAlaArgSerAsnValGluLysSerLysLysThrLeuGluSerGlnLeuVal 1097

Qy        520 ------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA--------- 564
Db       1098 AlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAlaLeuGluLysLysLys 1117

Qy        564 ------------------------------------------------------------ 564

Db       1118 LysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGluSerThrGlyGlyGlu 1137

Qy        565 --------------------------GAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db       1138 LysLysSerLeuTyrAspLeuLysValLysGlnGluSerAspMetGluAlaLeuArgAsn 1157

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGAT------ 651
Db       1158 GlnIleSerGluLeuGlnSerThrIleAlaLysLeuGluLysIleLysSerThrLeuGlu 1177

Qy        652 ---------------------------------------------ACGAAAAAAAATTTA 666
Db       1178 GlyGluValAlaArgLeuGlnGlyGluLeuGluAlaGluGlnLeuAlaLysSerAsnVal 1197

Qy        667 GAAAGAAAAAAGGAACATGGAGATATA----------------------TTAGCAGAG 702
Db       1198 GluLysGlnLysLysLysValGluLeuAspLeuGluAspLysSerAlaGlnLeuAlaGlu 1217

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
Db       1218 GluThrAlaAlaLysGlnAlaLeuAspLysLeuLysLysLysLeuGluGlnGluLeuSer 1237

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC 822
Db       1238 GluVal---------GlnThrGlnLeuSerGluAlaAsnAsnLysAsnValAsnSerAsp 1254

Qy        823 AAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGA 882
Db       1255 SerThrAsnLysHisLeuGlu---ThrSerPheAsnAsnLeuLysLeuGluLeuGluAla 1273

Qy        883 CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--------------- 927
Db       1274 GluGlnLysAla---LysGlnAlaLeuGluLysLysArgLeuGlyLeuGluSerGluLeu 1292

Qy        928 --------------TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAA 972
Db       1293 LysHisValAsnGluGlnLeuGluGluGluLysLysGlnLysGluSerAsnGluLysArg 1312
```

```
Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
              ||||||         :::|||::::::   |||          :::
Db       1313 LysValAspLeuGluLys---GluValSerGluLeuLysAsp-----------GlnIle 1327

Qy       1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------ 1086
              |||:::|||:::::::::::          :::   |||   ::::::
Db       1328 GluGluGluValAlaSerLys---------LysAlaValThrGluAlaLysAsnLysLys 1344

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
              :::   :::|||||||||   |||         ||||||   :::
Db       1345 GluSerGluLeuAspGluIleLysArg------GlnTyrAlaAspValValSerSerArg 1362

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT 1206
              :::      :::   ::::::|||:::   |||      ||||||||   |||
Db       1363 AspLysSerVal---GluGlnLeuLysThrLeuGlnAlaLysAsnGluGluLeu------ 1379

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
                          :::::::::   |||   :::|||   ::::::   :::
Db       1380 -----------------ArgAsnThrAlaGluGluAlaGluGlyGlnLeuAspArgAla 1393

Qy       1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
              ::::::   :::|||      :::   :::   :::         |||:::|||   :::
Db       1394 GluArgSerLysLysLysAlaGluPheAspLeuGluGluAlaValLysAsnLeuGluGlu 1413

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
              |||      ||||||   |||   :::|||   :::   |||   :::      :::
Db       1414 GluThrAlaLysLysValLysAlaGluLysAlaMetLysLysAlaGluThrAspTyrArg 1433

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAAT--------------------GAAATT 1425
                 :::         :::   |||      |||                   :::|||
Db       1434 SerThrLysSerGluLeuAspAspAlaLysAsnValSerSerGluGlnTyrValGlnIle 1453

Qy       1426 TTACAGATCGTGGATGAGTTATCTGAA 1452
                 ::::::   :::||||||||||||
Db       1454 LysArgLeuAsnGluGluLeuSerGlu 1462
```

RESULT 29
T16416
hypothetical protein F52B10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T16416
R;Wilcox, L.
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of C. elegans cosmid F52B10.
A;Reference number: Z18510
A;Accession: T16416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1956 <WIL>
A;Cross-references: UNIPROT:Q20641; UNIPARC:UPI0000177613; EMBL:U41990; NID:g1118154; PI
C;Genetics:
A;Gene: CESP:F52B10.1
A;Introns: 44/2; 108/3; 161/1; 211/3; 242/3; 297/1; 345/1; 503/1; 876/3; 941/3; 1134/2;
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;81-763/Domain: myosin motor domain homology <MMO>

Alignment Scores:

Pred. No.:	1.64e-09	Length:	1956
Score:	309.50	Matches:	142
Percent Similarity:	41.06%	Conservative:	122
Best Local Similarity:	22.08%	Mismatches:	204
Query Match:	12.45%	Indels:	175
DB:	2	Gaps:	26

US-09-837-344-41 (1-1482) x T16416 (1-1956)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::   :::      ||||||||||||   :::|||::::::||||||
Db        945 ArgAspLeuGluGluGlnLeuGluGlnGluGluGlnAlaArgGlnLysLeuLeuAsp 964

Qy         61 CAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 117
                                                  :::|||      |||      :::|||
Db        965 LysThrAsnValAspGlnArgLeuArgAsnLeuGluGluArgLeuValGluLeuGlnAsp 984

Qy        118 GAT------------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
                            |||:::::::::||||||   :::||||||::::::
Db        985 AlaTyrAspLysLeuLeuLysGluLysArgLeuGluGluLysValGluGlyLeuThr 1004

Qy        166 AGCGAT---CTAGAACAAGAGAGACGTGCTAAA------------------------- 195
              :::      |||:::   |||   |||||||||
Db       1005 ThrGlnLeuLeuAspHisGluGluArgAlaLysHisGlyValLysAlaLysGlyArgLeu 1024

Qy        196 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
                 :::|||   |||   :::   ||||||   :::||||||:::   |||   :::
Db       1025 GluAsnGlnLeuHisGluLeuGluGlnAspLeuAsnArgGluArgGlnTyrLysSerGlu 1044

Qy        253 TTGCAAGAACAACAAAGCGATTTA-----------------GAACAAGATAGACTTGCT 294
              |||::::::   :::      |||                  :::|||   ||||||
Db       1045 LeuGluGlnHisLysArgLysLeuLeuAlaGluLeuGluAspSerLysAspHisLeuAla 1064

Qy        295 AAA-------------------------------------------------GAAAAGTTA 306
              :::                                                 |||:::|||
Db       1065 GluLysMetGlyLysValGluGluLeuAsnAsnGlnLeuMetLysArgAspGluGluLeu 1084

Qy        307 CAAGAGCAG---------------------------------------------------- 315
              |||   |||
Db       1085 GlnHisGlnLeuThrArgTyrAspGluGluSerAlaAsnValThrLeuMetGlnLysGln 1104

Qy        316 --------------------------------CAAAGCGATTTAGAACAAGAGAGACGT 342
                                             :::   |||:::|||   ||||||
Db       1105 MetArgAspMetGlnThrThrIleAspGluLeuArgGluAspMetGluThrGluArgAsn 1124

Qy        343 GCTAAAGAAAAG------------------------------------------------ 354
              |||:::   |||
Db       1125 AlaArgAsnLysAlaGluMetThrArgArgGluValValAlaGlnLeuGluLysValLys 1144

Qy        355 ----------------------------------TTGCAAGAACAACAAAGCGATTTA 378
                                               ||||||:::      |||
Db       1145 GlyAspValLeuAspLysValAspGluAlaThrMetLeuGlnAspLeuMetSerArgLys 1164

Qy        379 GAACAAGAG------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 426
              ::::::|||           :::||||||   |||::::::|||
Db       1165 AspGluGluValAsnAlaThrLysArgAlaIleGluGlnIleGlnHisThrMetGluGly 1184

Qy        427 TTAGAACAAGAGAGACTTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::|||:::   ||||||          :::::::::|||                :::
Db       1185 LysIleGluGluGlnLysAlaLysPheSerArgGlnValGluGluLeuHisAspGlnIle 1204

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||   ::::::   :::   :::|||::::::||||||:::      :::||||||
Db       1205 GluGlnHisLysLysGlnArgSerGlnLeuGluLysGlnGlnAsnGlnAlaAspGlnGlu 1224

Qy        541 AGA------CGTGCTAAAGAA------AAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
              |||           |||:::|||          ||||||      :::::::|||:::::::::
Db       1225 ArgAlaAspMetAlaGlnGluIleAlaLeuLeuGlnAlaSerArgAlaAspIleAspLys 1244

Qy        589 GAGAGACGTGCTAAAGAA---AAGTTGCAAGAGCAGCAAAGAGATTTAGAA--------- 636
              :::|||:::      |||      |||   |||   |||   :::|||
Db       1245 LysArgLysIleHisGluAlaHisLeuMetGluIleGlnAlaAsnLeuAlaGluSerAsp 1264

Qy        637 ----------------------------CAAAGGAAGGCTGATACGAAAAAAAATTTA 666
                                           :::|||      |||         :::|||
Db       1265 GluHisLysArgThrLeuIleAspGlnLeuGluArgSerArgAspGluLeuAspHisLeu 1284

Qy        667 GAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATA 726
                 |||   :::|||      :::                :::::      ||||||
Db       1285 AsnArgValArgGluGluGluGluHisAlaPheAlaAsnMetGlnArgArgLeuAlaThr 1304

Qy        727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786
                       ::::::          ||||||:::          |||      ::::::
Db       1305 AlaGluGlyGlnIleGlnGluLeuAsnGluGln---------IleGlnGluGluThrArg 1321

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGA---------GATTCCAAGGAAATATCTATA 837
              |||         |||         :::|||         |||   |||         :::
```

```
Db       1322 LeuLysIleAlaAsnIleAsnArgAlaArgGlnLeuGluAspGluLysAsnAlaLeuLeu 1341

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              ||||||          |||   :::      :::::::|||      :::::::||||||
Db       1342 AspGluLysGluGluAlaGluGlyLeuArgAlaHisLeuGlu------LysGluIleHis 1359

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                              :::   |||:::   :::       :::|||   ::::::
Db       1360 AlaAla--------------ArgGlnGlyAlaGlyGluAlaArgArgLysAlaGluGlu 1374

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
              |||   :::   |||          |||          |||:::   ||||||
Db       1375 SerValAsnGlnGlnLeuGluGluLeuArgLysLysAsnLeuArgAspValGluHisLeu 1394

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA---------TATGACGATTCATTA 1068
              |||          |||   |||::::::   |||
Db       1395 GlnLysGlnLeuGluGluSerGluValAlaLysGluArgIleLeuGlnSerLysLysLys 1414

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTA---------------------------- 1098
              |||   :::|||   :::|||      :::
Db       1415 IleGlnGlnGluLeuGluAspSerSerMetGluLeuGluAsnValArgAlaSerHisArg 1434

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
              |||      |||          ::::::::::::          ||||||   :::   :::
Db       1435 AspSerGluLysArgGlnLysLysPheGluSerGlnMetAlaGluGluArgValAlaVal 1454

Qy       1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------ 1206
                 |||   |||   |||          ::::::   |||   |||
Db       1455 GlnLysAlaLeuLeuAspArgAspAlaMetSerGlnGluLeuArgAspArgGluThrArg 1474

Qy       1207 ------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                        |||:::|||   ::::::          |||   |||   |||
Db       1475 ValLeuSerLeuLeuAsnGlu---ValAspIleMetLysGluHisLeuGluGluSerAsp 1493

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
                    :::            ::::::   :::|||:::            :::|||
Db       1494 ArgValArgArgSerLeuGlnGlnGluLeuGlnAspSerIleSerAsnLysAspAspPhe 1513

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAA--------------------AATGAT 1353
                           ::::::|||   |||                        ||||||
Db       1514 GlyLysAsnValHisGluLeuGluLysAlaLysArgSerLeuGluAlaGluLeuAsnAsp 1533

Qy       1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA 1413
                 :::|||      |||:::   :::
Db       1534 MetArgValGlnMetGluGluLeuGlu-------------------------------- 1542

Qy       1414 GACAATGAAATTTTACAGATCGTGGATGAG------TTATCTGAAGATATAACTAAATAT 1467
              ||||||      |||||||||   ::::::      |||      ::::::|||
Db       1543 AspAsn------LeuGlnIleAlaGluAspAlaArgLeuArgLeuGluValThrAsnGln 1560

Qy       1468 TTTATGAAA 1476
                 :::|||
Db       1561 AlaLeuLys 1563

RESULT 30
A59234
slow myosin heavy chain 3 - quail
C;Species: Coturnix coturnix
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C;Accession: A59234
R;Nikovits Jr., W.; Wang, G.F.; Feldman, J.L.; Miller, J.B.; Wade, R.; Nelson, L.; Stock
J. Biol. Chem. 271, 17047-17056, 1996
A;Title: Isolation and characterization of an avian slow myosin heavy chain gene expres
A;Reference number: A59234; MUID:96291845; PMID:8663323
A;Accession: A59234
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1931 <NIK>
A;Cross-references: UNIPARC:UPI00000FB7EB; GB:U53862; NID:g1289513; PIDN:AAC59912.1; PID
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;81-761/Domain: myosin motor domain homology <MMO>

Alignment Scores:

Pred. No.:              1.75e-09        Length:         1931
Score:                  309.00          Matches:        153
Percent Similarity:     41.79%          Conservative:   127
Best Local Similarity:  22.84%          Mismatches:     192
Query Match:            12.43%          Indels:         198
DB:                     2               Gaps:           33

US-09-837-344-41 (1-1482) x A59234 (1-1931)

Qy           4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
               |||::::::   ::::::      |||      |||||||::   ||||||:::::::::
Db         922 GluGluGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 941

Qy          61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                 |||:::|||::::::|||                                ::::::   |||
Db         942 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuSerLeuAlaLysValGluLys 961

Qy          94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
               ||||||                          |||   |||:::   :::   |||:::
Db         962 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 981

Qy         127 CAAGAG------AGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
               :::         :::|||   |||||||||      |||||||||      ||||||:::
Db         982 GluAsnIleThrLysLeuThrLysGluLysLysThrLeuGlnGluSerHisGlnGlnAla 1001

Qy         169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                            :::||||||   |||
Db        1002 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuAlaLysAlaLysValLys 1021

Qy         202 TTGCAAGAACAACAAAGCGATTTAGAA-----------CAAGAGAGACGTGCTAAAGAA 249
               |||::::::|||      |||||||||                ||||||::::::   :::
Db        1022 LeuGluGlnGlnAlaAspAspLeuGluSerSerLeuGlnGlnGluLysLysIleArgMet 1041

Qy         250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
               |||:::      :::      ||||||   |||   :::||||||:::|||   :::
Db        1042 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerVal 1061

Qy         307 CAAGAGCAGCAAAGCGATTTAGAACAA---GAGAGACGTGCTAAAGAAAAG--------- 354
                  :::   ::::::|||   :::|||   |||   |||   |||:::|||
Db        1062 MetAspLeuGluAsnAspLysGlnGlnLeuGluGluArgLeuLysLysLysAspPheGlu 1081

Qy         355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
               |||              :::   :::|||   |||:::              :::||||||:::
Db        1082 LeuAsnThrLeuAsnAlaArgIleGluAspGluGlnAlaIleAlaAlaGlnLeuGlnLys 1101

Qy         415 CAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
               :::      :::|||:::                    :::::::::|||   :::
Db        1102 LysLeuLysGluLeuGln-----------------AlaArgIleGluGluLeuGluGlu 1115

Qy         475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
               :::||||||   ||||||      :::   |||:::::::::   :::|||:::|||
Db        1116 GluLeuGluAlaGluArgThrGlyArgAlaLysValGluLysLeuArgSerGluLeuLeu 1135

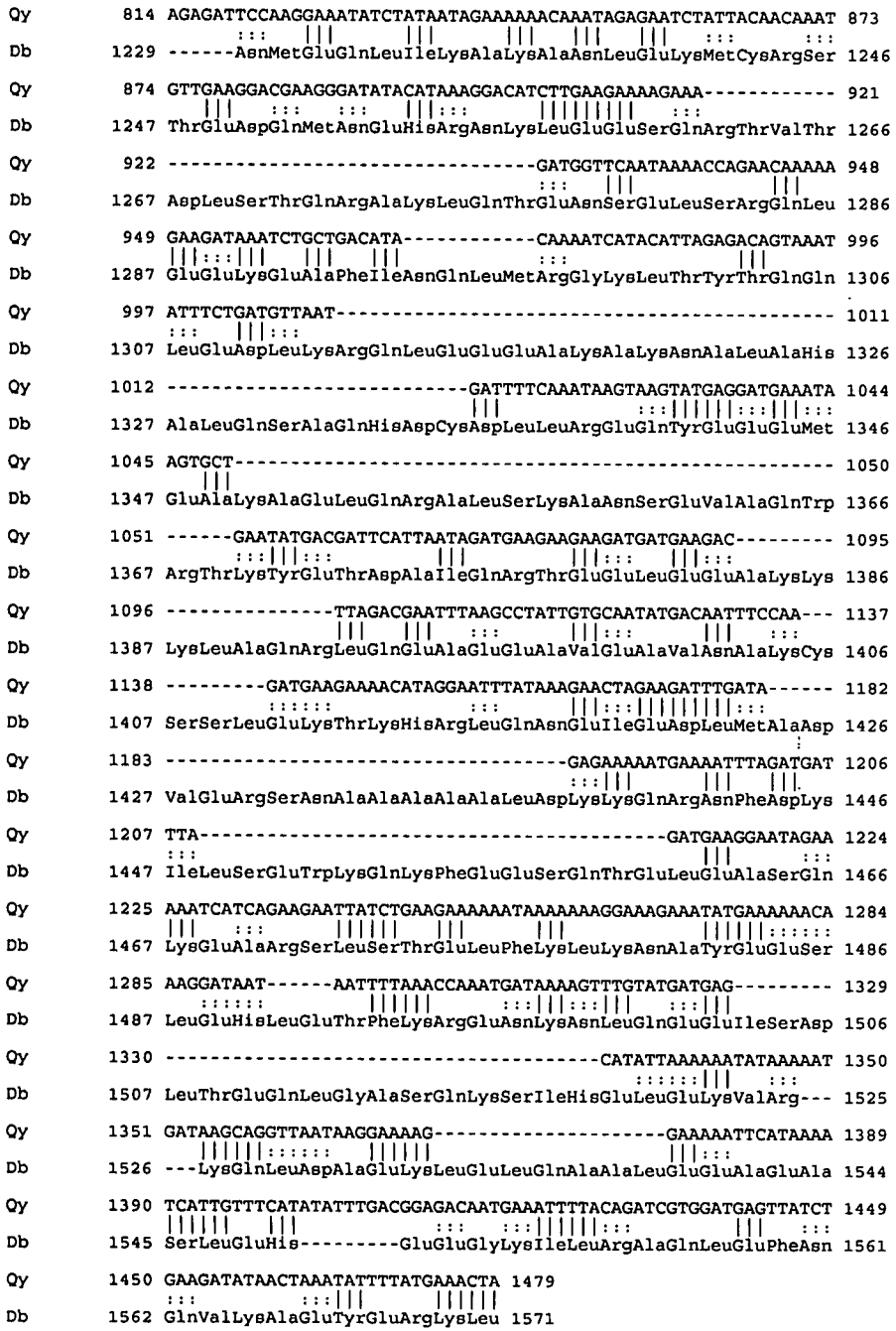
Qy         535 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAA-----------CAACAAAGCGAT 579
               ||||||                ||||:::|||:::|||
Db        1136 GlnGluLeuGluGluThrSerGluArgLeuGluGluAlaGlyGlyAlaThrSerValGln 1155

Qy         580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
               ||||||      ::::::   :::   :::   |||:::   :::||||||||||||:::
Db        1156 LeuGluLeuAsnLysLysArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGlu 1175

Qy         640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACATGGAGATATATTA 696
                           :::                  :::|||:::|||   |||   :::
Db        1176 AlaThrLeuGlnHisGluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerVal 1195

Qy         697 GCAGAGGATTTATATGGTCGTTTAGAA---ATACCAGCTATAGAACTTCCATCAGAAAAT 753
               ||||||   |||      :::|||:::   :::      ::::::         |||
Db        1196 AlaGlu---LeuSerGluGlnLeuAspAsnMetGlnArgValLysGlnLysLeuGluLys 1214

Qy         754 GAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT 813
               |||:::                ::::      |||      |||:::      :::
Db        1215 GluLysSer--------------GluLeuLysLeuGluLeuAspAspValSerSer--- 1228
```



Search completed: December 21, 2005, 11:31:01
Job time : 159.302 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:09:34 ; Search time 7876.23 Seconds
(without alignments)
8803.514 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	291.4	19.7	317	10	T09652	T09652 0224m3 gmbP
	2	256	17.3	1721	8	DR145459	DR145459 49056454
	3	249.4	16.8	1808	8	DR125630	DR125630 49096520
c	4	248.4	16.8	1519	10	AG386893	AG386893 Mus muscu
c	5	247.2	16.7	1453	10	AJ591978	AJ591978 Arabidops
	6	244.4	16.5	2071	8	DR147304	DR147304 49027161
	7	243	16.4	1932	8	DR141720	DR141720 49300102
	8	241	16.3	1727	10	CL114085	CL114085 ISB1-59J1
c	9	240	16.2	1491	10	CG753221	CG753221 P048-2-A0
	10	238.4	16.1	1783	8	DR141924	DR141924 49190649
c	11	237.8	16.0	1434	10	AJ592058	AJ592058 Arabidops
	12	237.8	16.0	1808	10	CL078963	CL078963 CH216-154
	13	235	15.9	1654	10	CL118709	CL118709 ISB1-72I1
	14	234.2	15.8	1797	10	CL066150	CL066150 CH216-108
	15	234	15.8	1829	10	CL109409	CL109409 ISB1-51H2
	16	233.8	15.8	1736	10	CL066373	CL066373 CH216-108
c	17	233.6	15.8	1407	10	AJ592026	AJ592026 Arabidops
	18	233.6	15.8	1753	8	DR131208	DR131208 49217594
	19	232.8	15.7	1512	10	CL113943	CL113943 ISB1-59E2
	20	232.8	15.7	1531	10	CG748014	CG748014 P041-4-B0
c	21	232.6	15.7	1811	10	CG753732	CG753732 P048-4-G0
	22	231.6	15.6	1917	10	CL081899	CL081899 CH216-164
	23	231.4	15.6	1650	10	CL041163	CL041163 CH216-51P
	24	231.2	15.6	1762	10	CL082587	CL082587 CH216-168
	25	230.4	15.5	1824	10	CL081234	CL081234 CH216-161
	26	230.2	15.5	1727	10	CL118815	CL118815 ISB1-72O5
	27	230	15.5	1632	10	CL082569	CL082569 CH216-167
	28	229.8	15.5	1616	10	CL081995	CL081995 CH216-165
	29	229.6	15.5	1630	10	CL078686	CL078686 CH216-152
	30	229.2	15.5	1457	10	CL082658	CL082658 CH216-169
c	31	228.8	15.4	1566	10	CG757757	CG757757 P053-1-D0
	32	228.6	15.4	1434	9	CC187638	CC187638 CH261-98P
c	33	228.4	15.4	1507	10	AG346189	AG346189 Mus muscu
	34	228.4	15.4	1823	10	CL058635	CL058635 CH216-88H
	35	228	15.4	1259	11	CR004722	CR004722 Forward s
	36	228	15.4	2033	10	CL081977	CL081977 CH216-165
c	37	228	15.4	2103	10	CL041093	CL041093 CH216-51M
	38	227.6	15.4	1712	10	CL078539	CL078539 CH216-151
	39	226.8	15.3	1547	10	CL066420	CL066420 CH216-108
	40	226.6	15.3	1377	10	CL077307	CL077307 CH216-144
	41	226.4	15.3	1716	9	CC222065	CC222065 CH261-11A
	42	225.6	15.2	1562	10	CL078533	CL078533 CH216-151
	43	225.6	15.2	1651	10	CL079030	CL079030 CH216-154
c	44	225.4	15.2	1482	10	AG332292	AG332292 Mus muscu
	45	224.2	15.1	1506	10	CL082761	CL082761 CH216-171

ALIGNMENTS

RESULT 1
T09652
LOCUS T09652 317 bp DNA linear GSS 05-DEC-2000
DEFINITION 0224m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic clone 0224m, genomic survey sequence.
ACCESSION T09652
VERSION T09652.1 GI:319484
KEYWORDS GSS.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 317)
AUTHORS Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C. and Dame,J.B.
TITLE Gene sequence tags from Plasmodium falciparum genomic DNA fragments prepared by the genease activity of mung bean nuclease
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
PUBMED 8234327
COMMENT Other_GSSs: 0224m7
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .317
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/db_xref="taxon:5833"
/clone="0224m"
/lab_host="E. coli XL1-Blue"
/clone_lib="gmbPfHB3.1, G. Roman Reddy"
/note="Vector: pBlueScript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were polished using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBlueScript SK(+). Recombinant plasmids

transformed E. coli XL1-Blue."
ORIGIN

Query Match 19.7%; Score 291.4; DB 10; Length 317;
Best Local Similarity 96.5%; Pred. No. 4.2e-38;
Matches 306; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```
Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||| |   ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 GAATTCGAAGCTATAGAACTTCCATCAGAGAATGAACGTGGATATTATATACCACATCAA 60

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 120

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 180

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        181 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGTACAAAAAGAAGATAAATCT 240

Qy        961 GCTGACATACAAAATCATACATTAGAGA-CAGTAAATATTTCTGATGTTAATGATTTTCA 1019
              |||||||||||||||||||||| |||  |||||| ||||||| ||||||||||||||||
Db        241 GCTGACATACAAAATCATACATTNGAGNCCAGTAANTATTTCTNATGTTAATGATTTTCA 300

Qy       1020 AATAAGTAAGTATGAGG 1036
              |||||||||||||||||
Db        301 AATAAGTAAGTATGAGG 317
```

RESULT 2
DR145459
LOCUS DR145459 1721 bp mRNA linear EST 16-JUN-2005
DEFINITION 49056454 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone A9 3', mRNA sequence.
ACCESSION DR145459
VERSION DR145459.1 GI:67890549
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1721)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226693903
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1721
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="A9"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

Query Match 17.3%; Score 256; DB 8; Length 1721;
Best Local Similarity 47.6%; Pred. No. 2.6e-32;
Matches 661; Conservative 0; Mismatches 727; Indels 0; Gaps 0;

```
Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
              || || |  ||| |     | |  || | | |       ||| ||||     || || ||
Db        137 AAAAAAANAAAAACCCAAAAAAGAAAAAAANANNAAAAAAAAAAAAAAAAAAAAAAAA 196

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
              |||   |   | || || | | |       ||| ||||   | || |  |  |||   |
Db        197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 256

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
              | || || | | |       ||| ||||     || || || |||    |   | || ||
Db        257 AAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAA 316

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
              | | |      |||  |||     || || || |||   |   | || || | | |
Db        317 AAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
              |||  ||||     || || || |||   |   | || || | | |       ||| |||
Db        377 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 436

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
              |   | || |  |  |||   |   | || || | | |       ||| ||||     ||
Db        437 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 496

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
              || || |||   |   | || || | | |       ||| ||||     || || || |||
Db        497 AAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 556

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
              |   | || || | | |       ||| ||||     || || || |||   |   |
Db        557 AAAANAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAANAAAAA 616

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
              || || | | |       ||| ||||     || || || |||   |   | || || | |
Db        617 AAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAA 676

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
              |      |||  ||||     || || || |       |   | || || | | |       |
Db        677 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNAAAANAAAAAAAAAAAAAAAAAAAAA 736

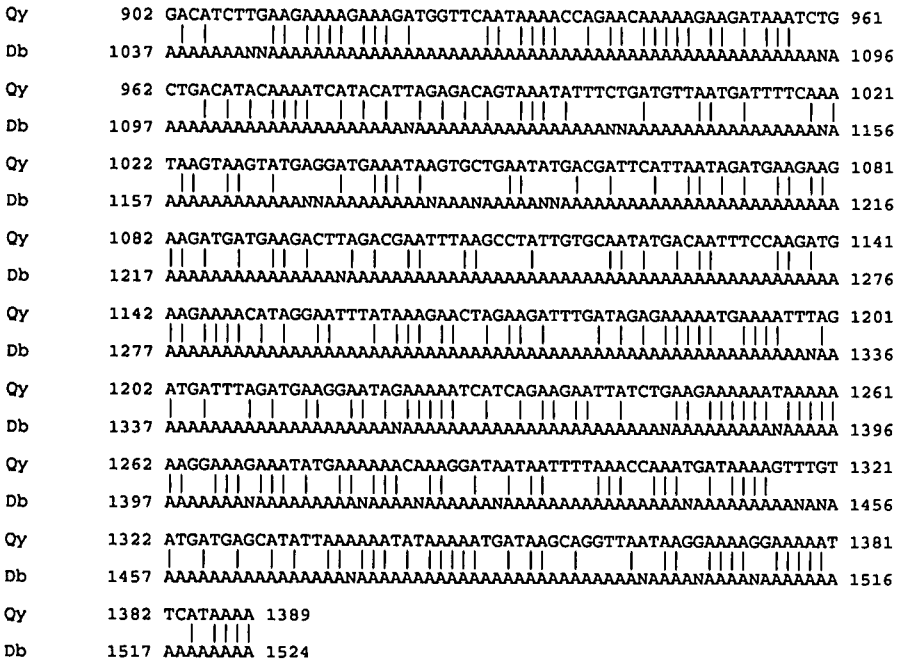
Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
              || ||||     || |  |  ||| | |   | || |||  ||      | |  |||||||
Db        737 AAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAA 796

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
              |   | ||| ||||||  || |     | | |  |  | | |   | |          |
Db        797 AAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 856

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
              ||| |  |  | | ||     | | ||||  ||    | |  | | | | |  ||
Db        857 AAAAAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAA 916

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAG 841
                    |   | | || | |   || |  | | |    ||  ||| |   | || |
Db        917 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 976

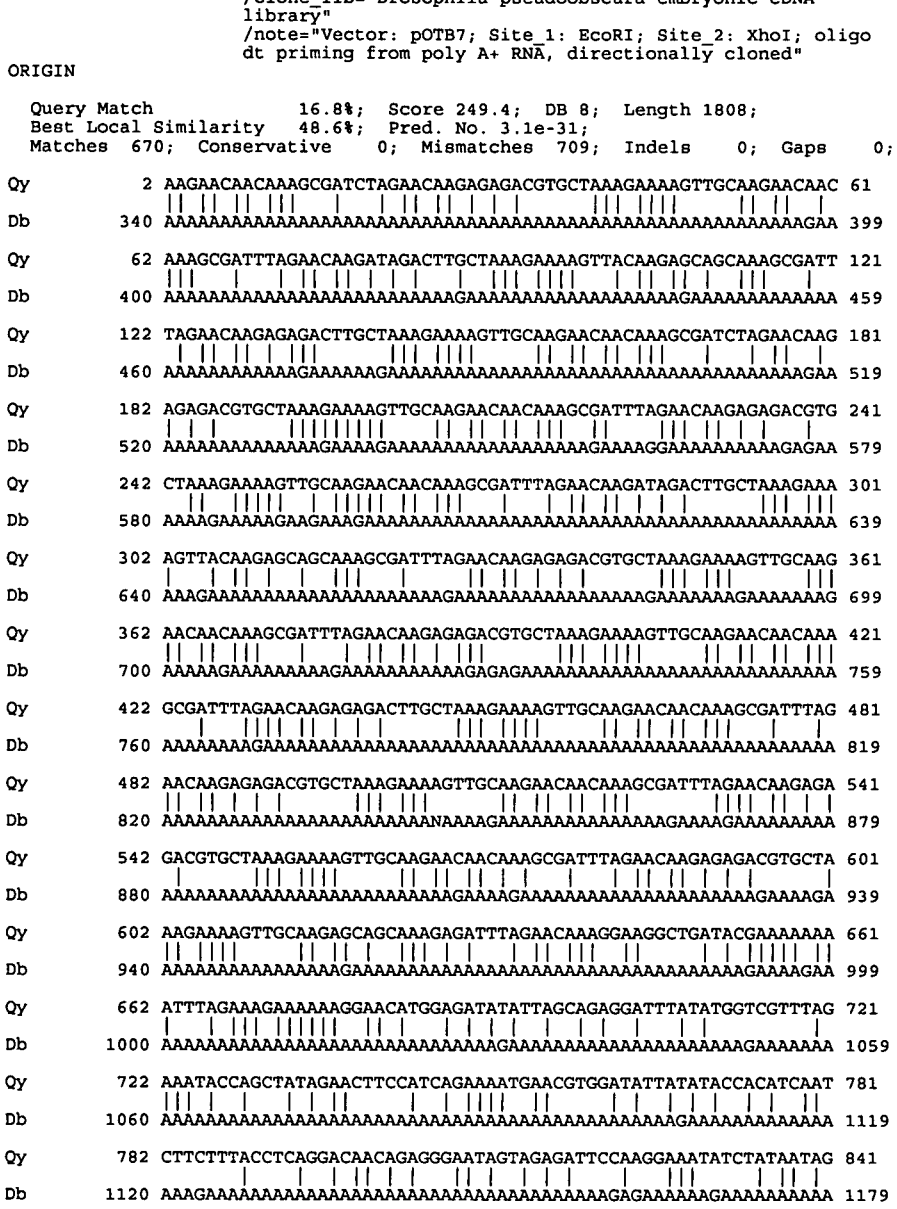
Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
              |||||| ||| | | ||   |  | || |||     ||  |  ||   | | | | |||
Db        977 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 1036
```

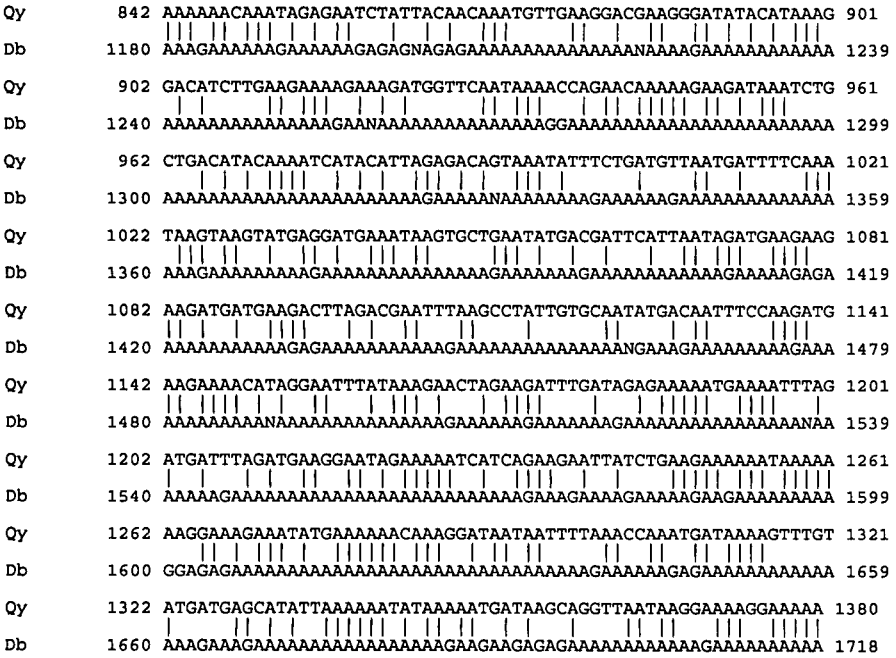


RESULT 3
DR125630
LOCUS DR125630 1808 bp mRNA linear EST 15-JUN-2005
DEFINITION 49096520 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O23 5', mRNA sequence.
ACCESSION DR125630
VERSION DR125630.1 GI:67842328
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226697953

Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1808
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O23"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

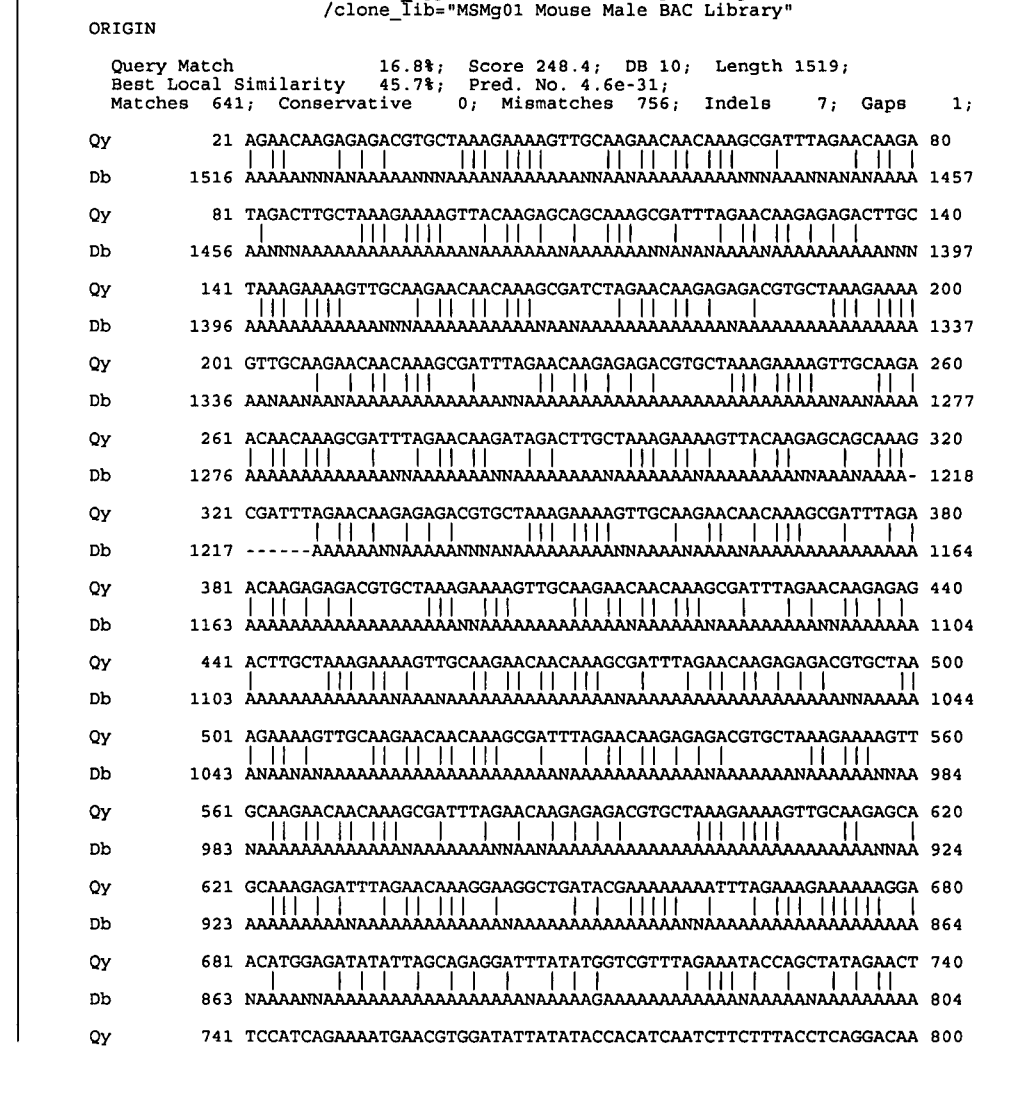
Query Match 16.8%; Score 249.4; DB 8; Length 1808;
Best Local Similarity 48.6%; Pred. No. 3.1e-31;
Matches 670; Conservative 0; Mismatches 709; Indels 0; Gaps 0;

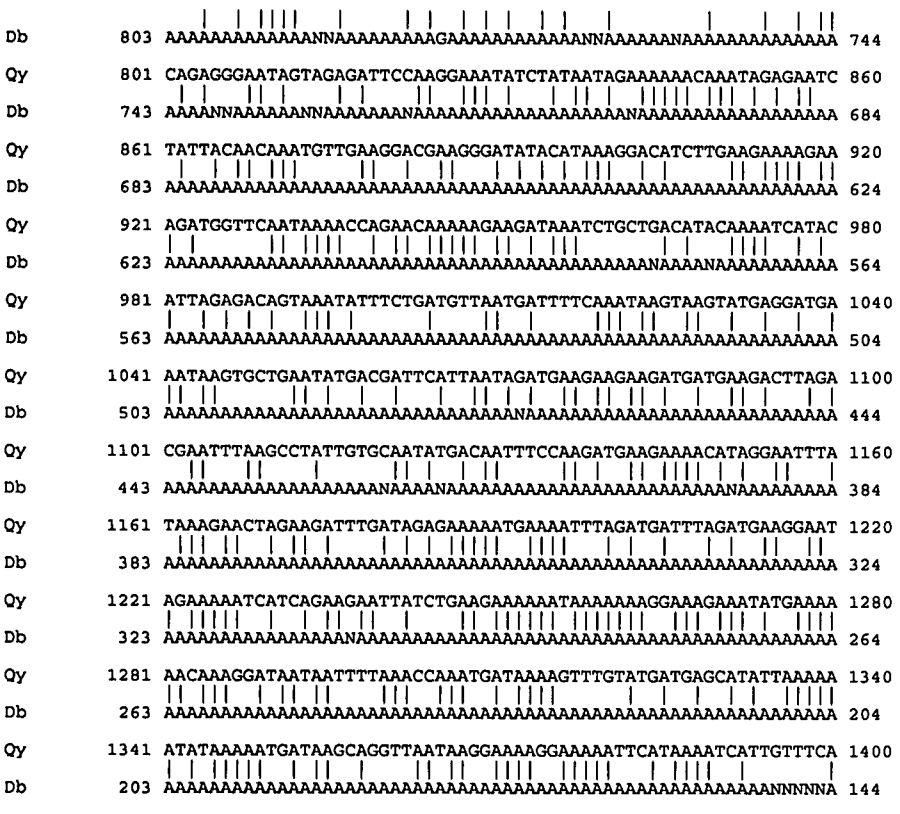




RESULT 4
AG386893/c
LOCUS AG386893 1519 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-201E07.TJ, genomic survey sequence.
ACCESSION AG386893
VERSION AG386893.1 GI:47998098
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1519)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1519
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-201E07.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

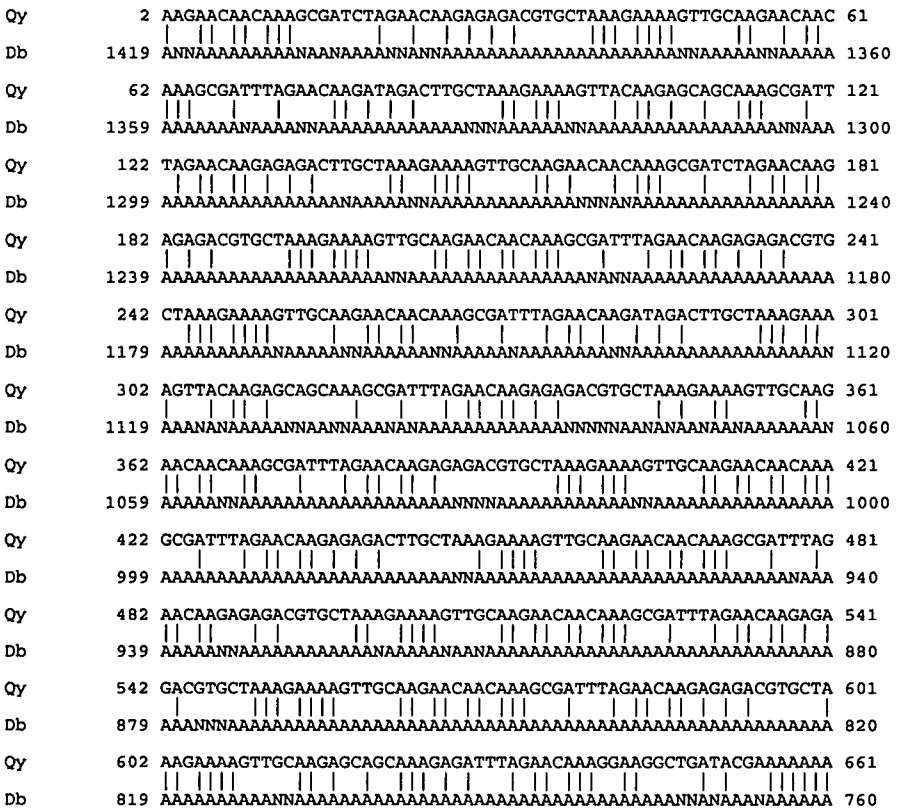
Query Match 16.8%; Score 248.4; DB 10; Length 1519;
Best Local Similarity 45.7%; Pred. No. 4.6e-31;
Matches 641; Conservative 0; Mismatches 756; Indels 7; Gaps 1;

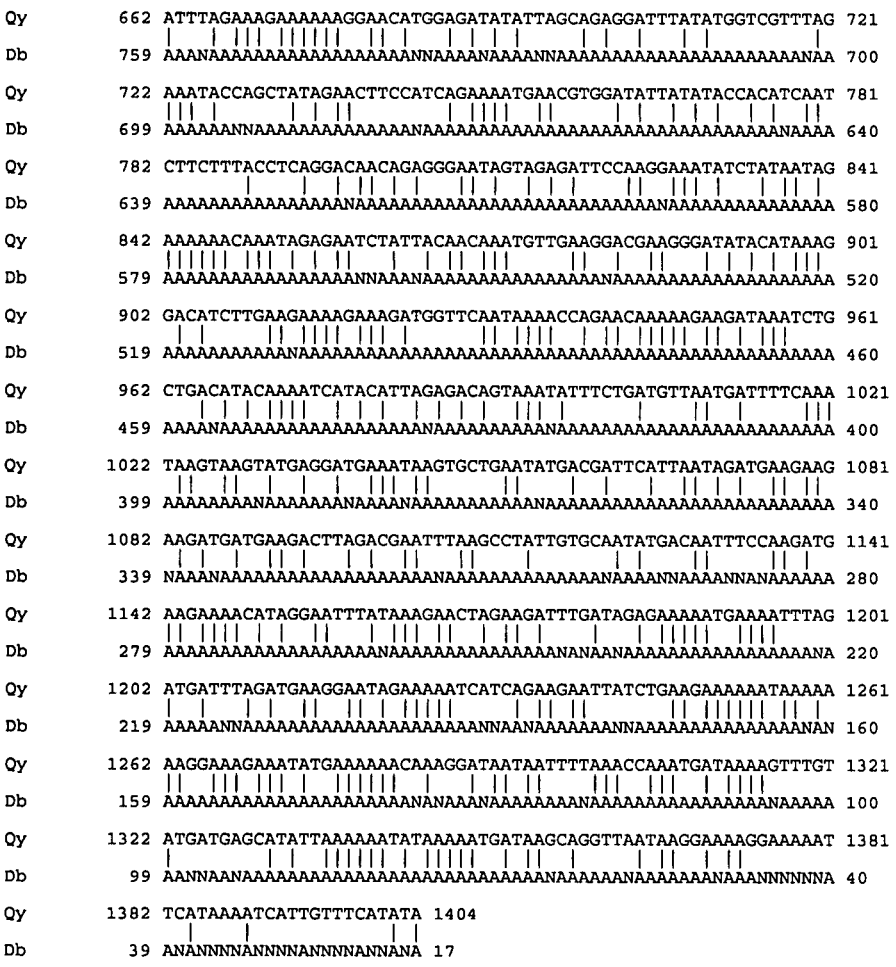




```
RESULT 5
AJ591978/c
LOCUS       AJ591978               1453 bp    DNA     linear   GSS 15-JAN-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            598D01, genomic survey sequence.
ACCESSION   AJ591978
VERSION     AJ591978.1  GI:37941602
KEYWORDS    GSS; right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 1453)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .1453
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="598D01"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .1453
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN
```

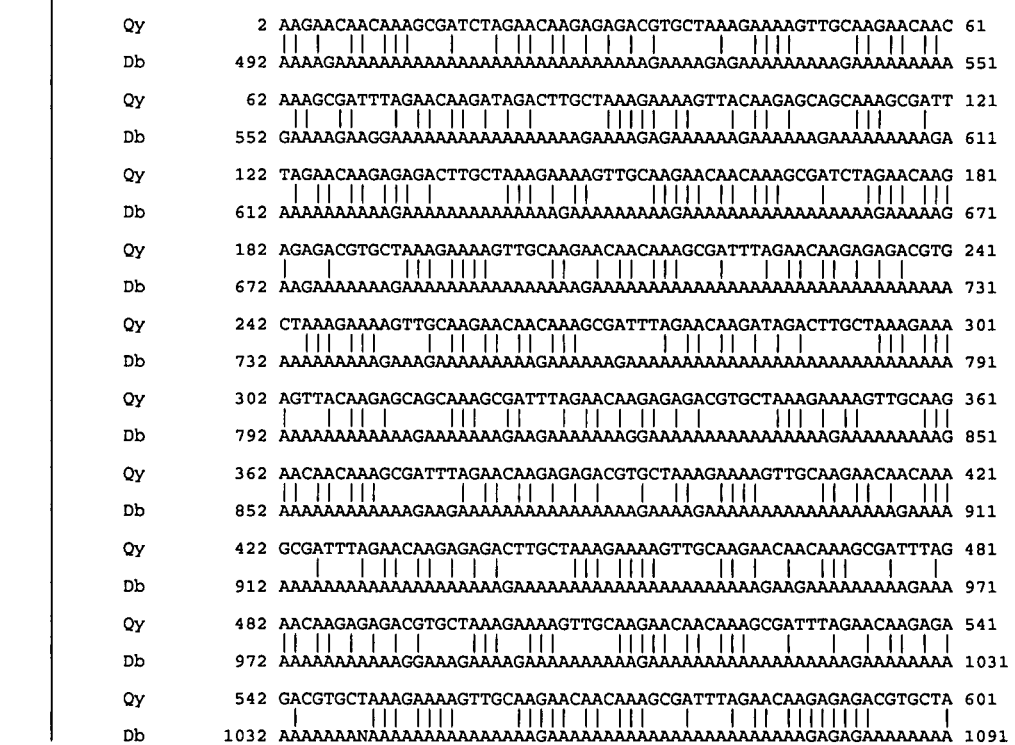
Query Match 16.7%; Score 247.2; DB 10; Length 1453;
Best Local Similarity 44.0%; Pred. No. 7.2e-31;
Matches 618; Conservative 0; Mismatches 785; Indels 0; Gaps 0;

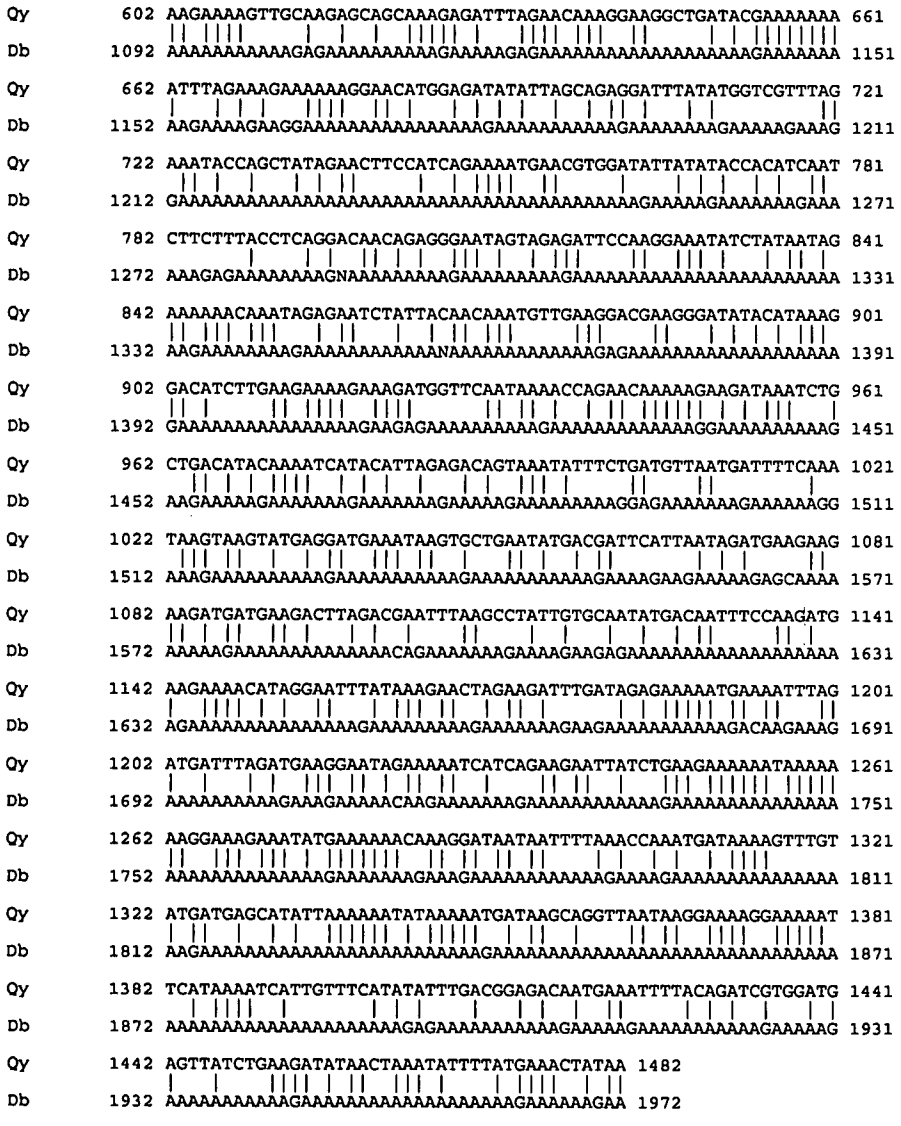




RESULT 6
DR147304
LOCUS DR147304 2071 bp mRNA linear EST 16-JUN-2005
DEFINITION 49027161 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O11 3', mRNA sequence.
ACCESSION DR147304
VERSION DR147304.1 GI:67892394
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2071)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226691688
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .2071
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O11"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

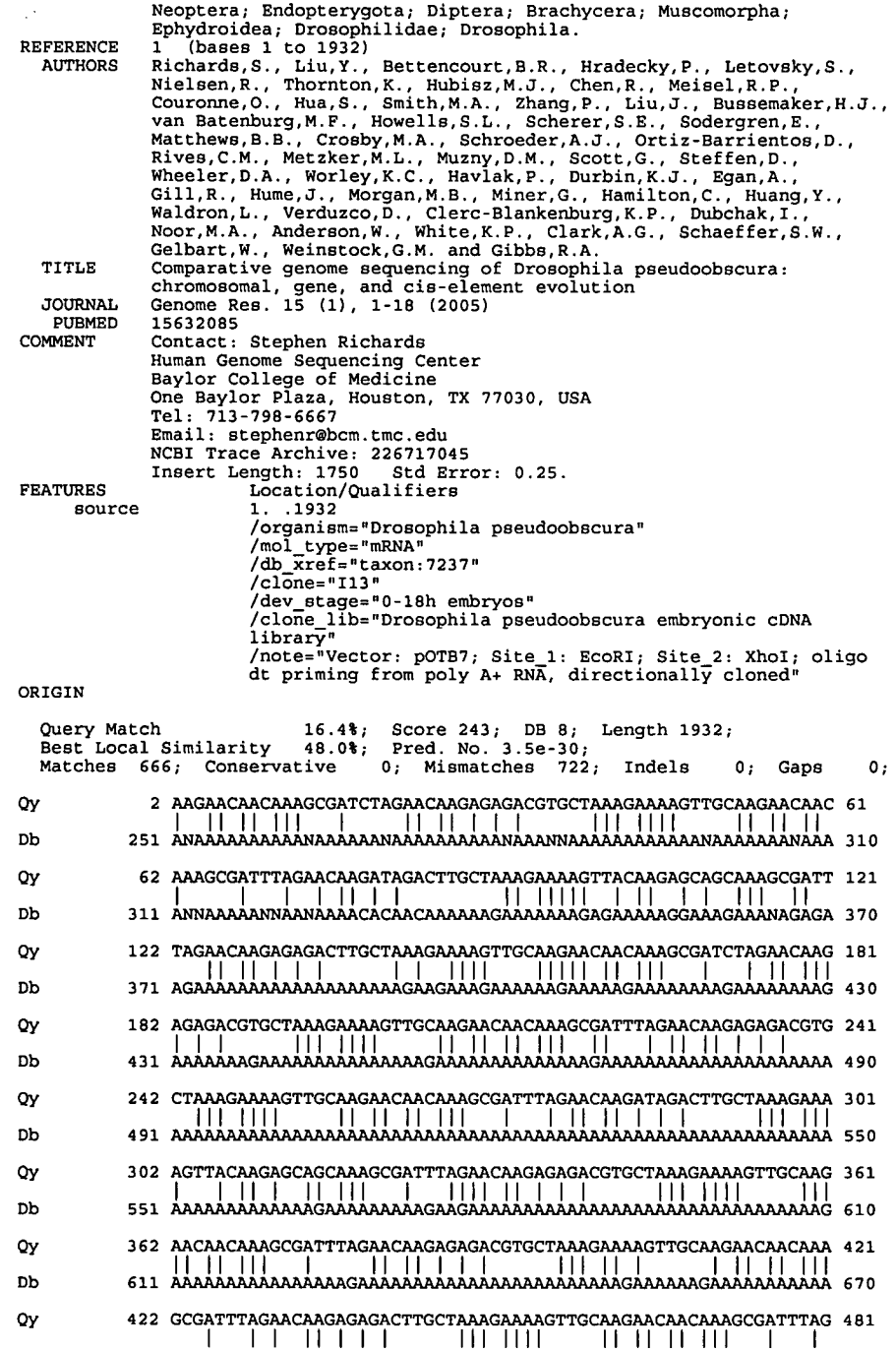
Query Match 16.5%; Score 244.4; DB 8; Length 2071;
Best Local Similarity 47.7%; Pred. No. 2.1e-30;
Matches 707; Conservative 0; Mismatches 774; Indels 0; Gaps 0;

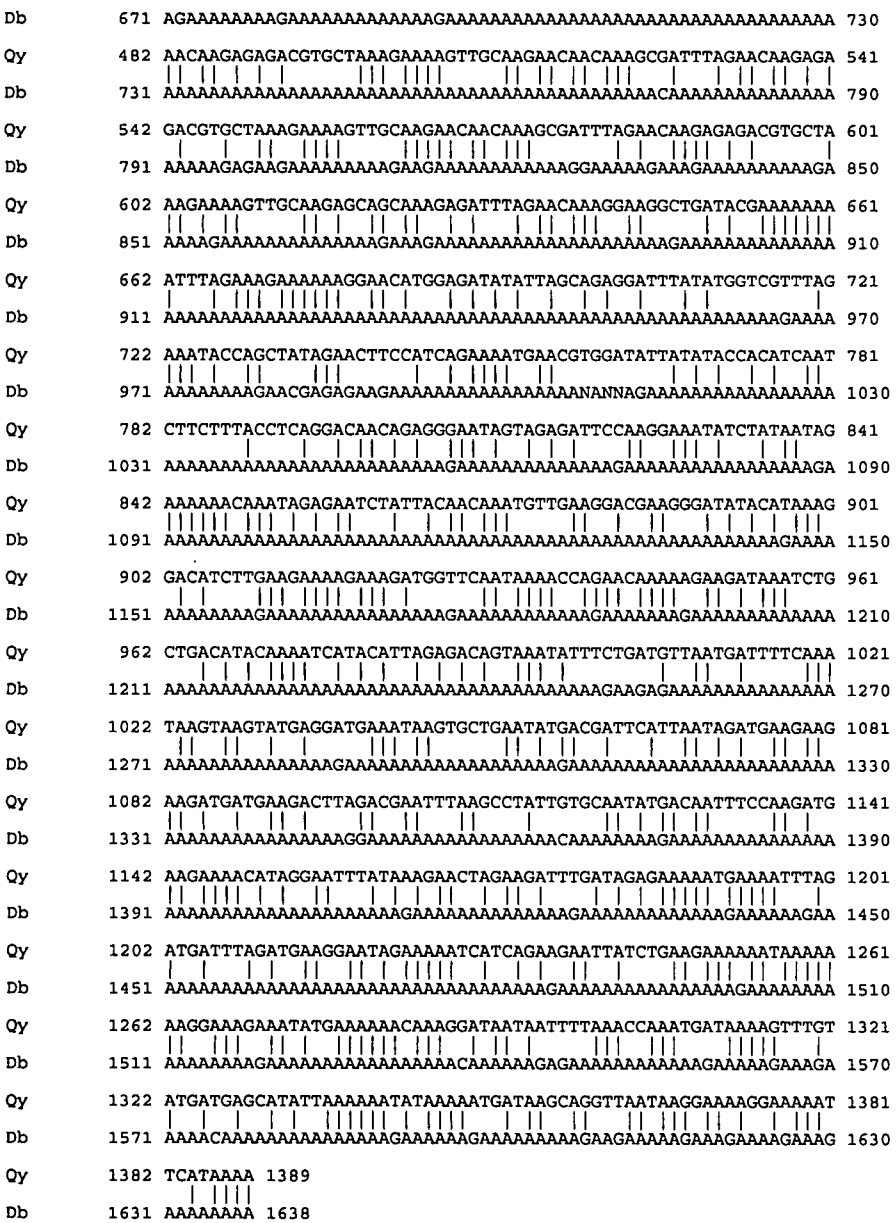




RESULT 7
DR141720
LOCUS DR141720 1932 bp mRNA linear EST 16-JUN-2005
DEFINITION 49300102 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone I13 3', mRNA sequence.
ACCESSION DR141720
VERSION DR141720.1 GI:67886810
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1932)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226717045
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1932
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="I13"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

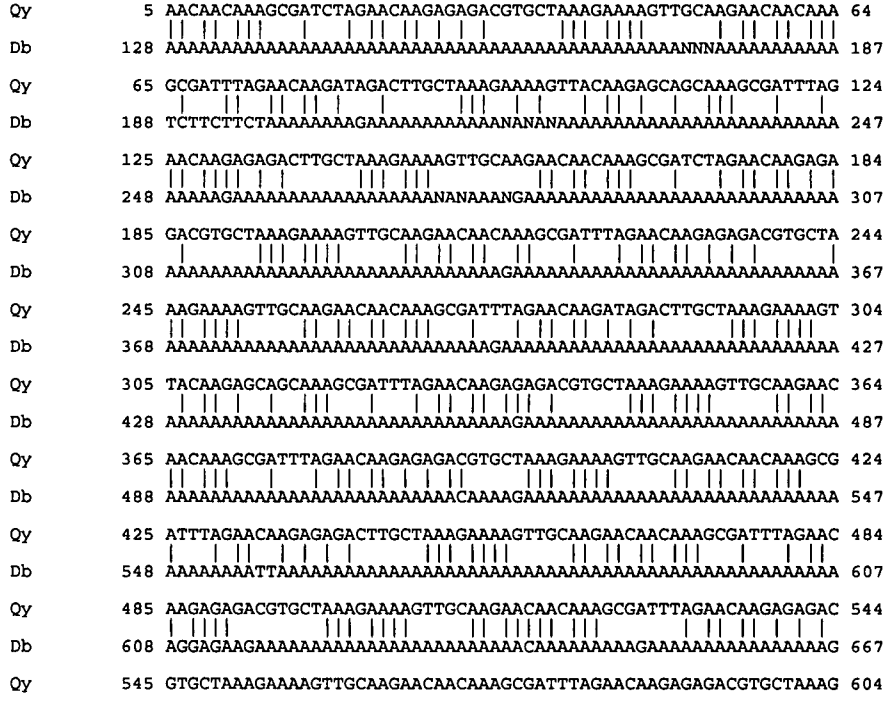
Query Match 16.4%; Score 243; DB 8; Length 1932;
Best Local Similarity 48.0%; Pred. No. 3.5e-30;
Matches 666; Conservative 0; Mismatches 722; Indels 0; Gaps 0;

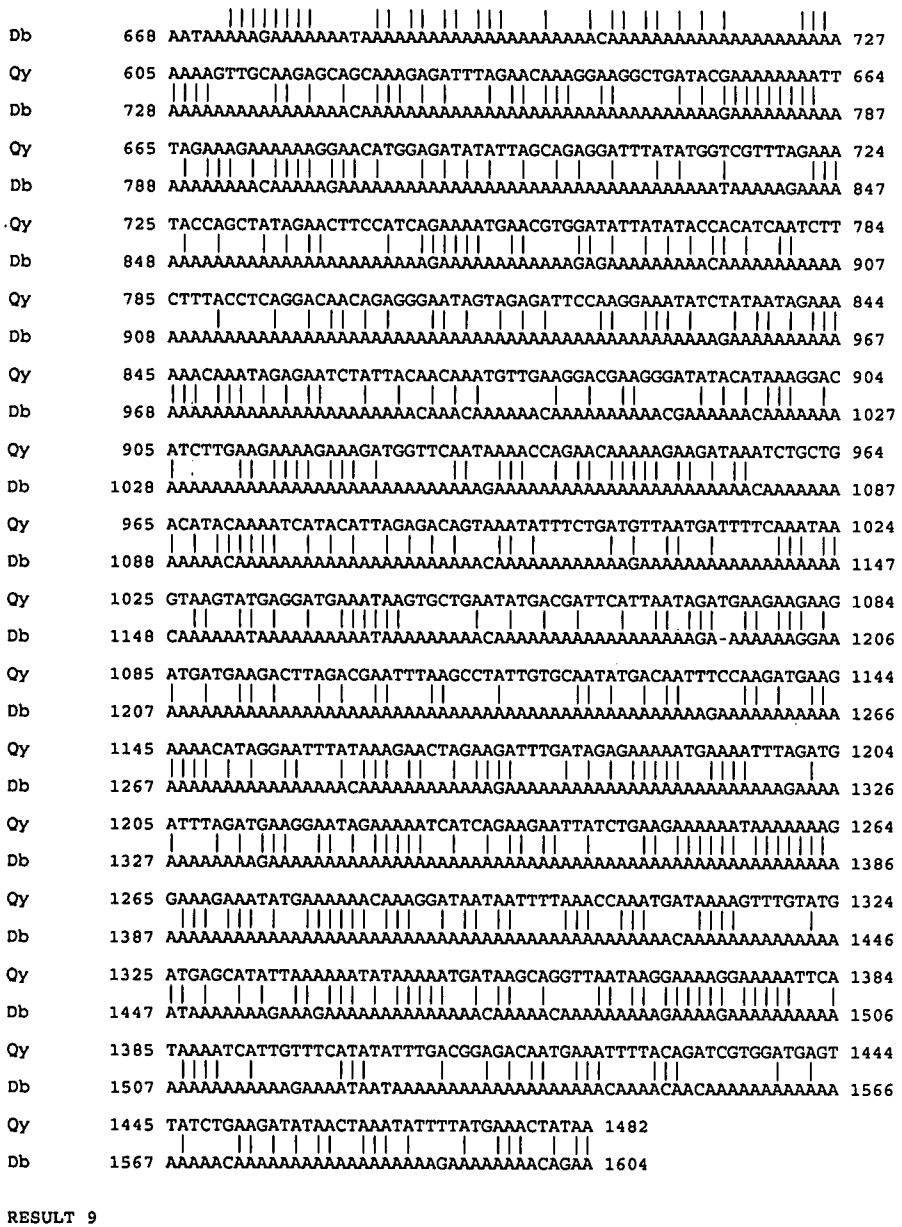




RESULT 8
CL114085
LOCUS CL114085 1727 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-59J11_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59J11, genomic survey sequence.
ACCESSION CL114085
VERSION CL114085.1 GI:40607720
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1727)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 280
High quality sequence stop: 1423.
FEATURES Location/Qualifiers
source 1. .1727
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-59J11"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

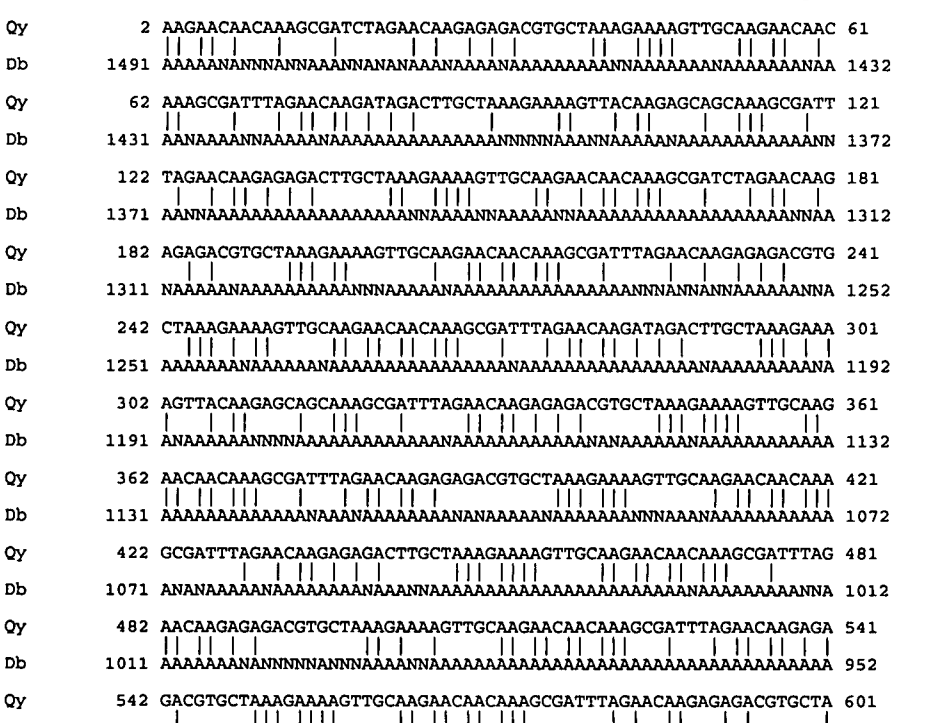
Query Match 16.3%; Score 241; DB 10; Length 1727;
Best Local Similarity 47.9%; Pred. No. 7.5e-30;
Matches 708; Conservative 0; Mismatches 769; Indels 1; Gaps 1;

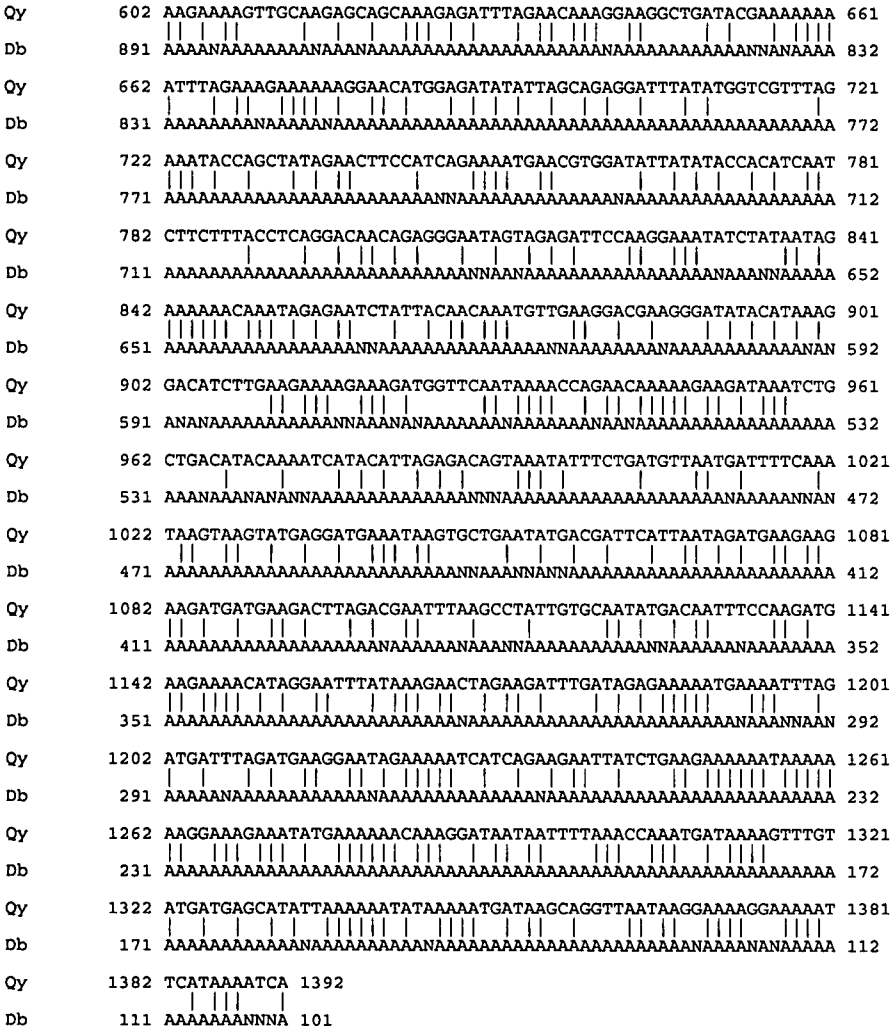




RESULT 9
CG753221/c
LOCUS CG753221 1491 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-2-A01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753221
VERSION CG753221.1 GI:37977480
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1491
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

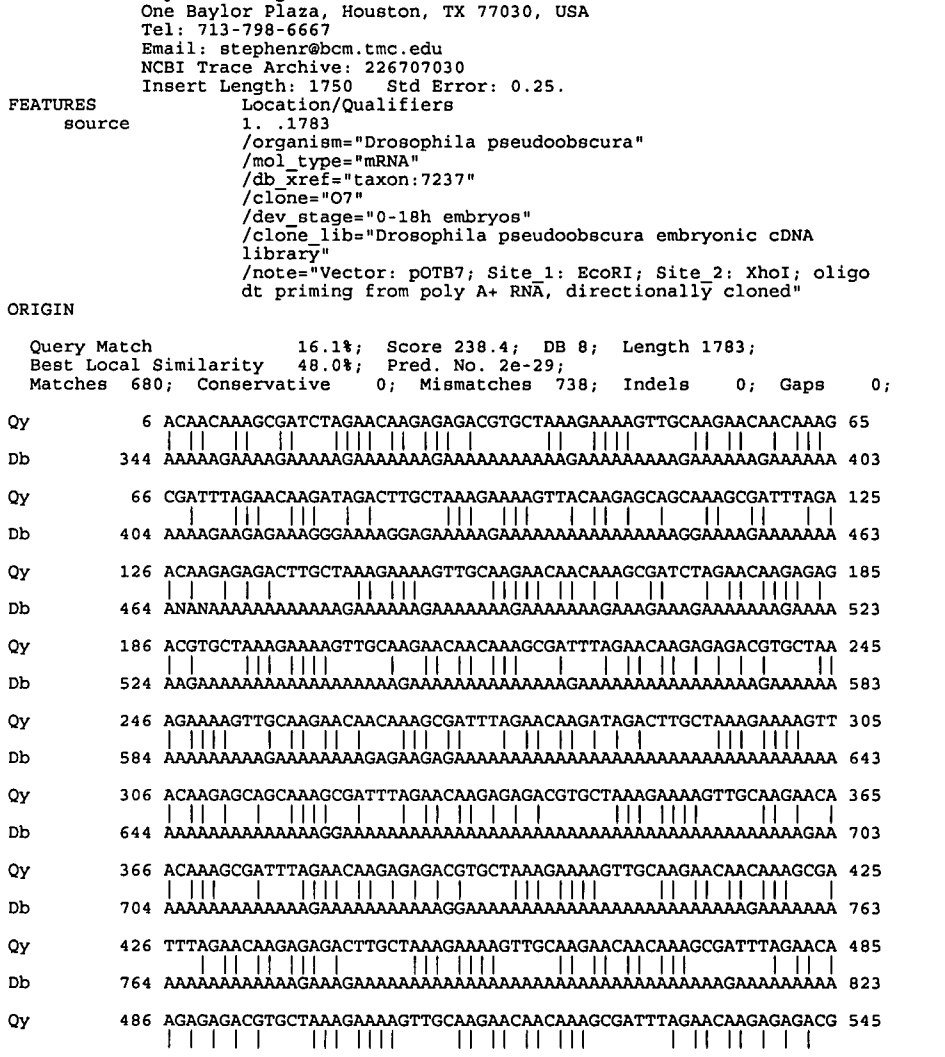
Query Match 16.2%; Score 240; DB 10; Length 1491;
Best Local Similarity 43.4%; Pred. No. 1.1e-29;
Matches 603; Conservative 0; Mismatches 788; Indels 0; Gaps 0;

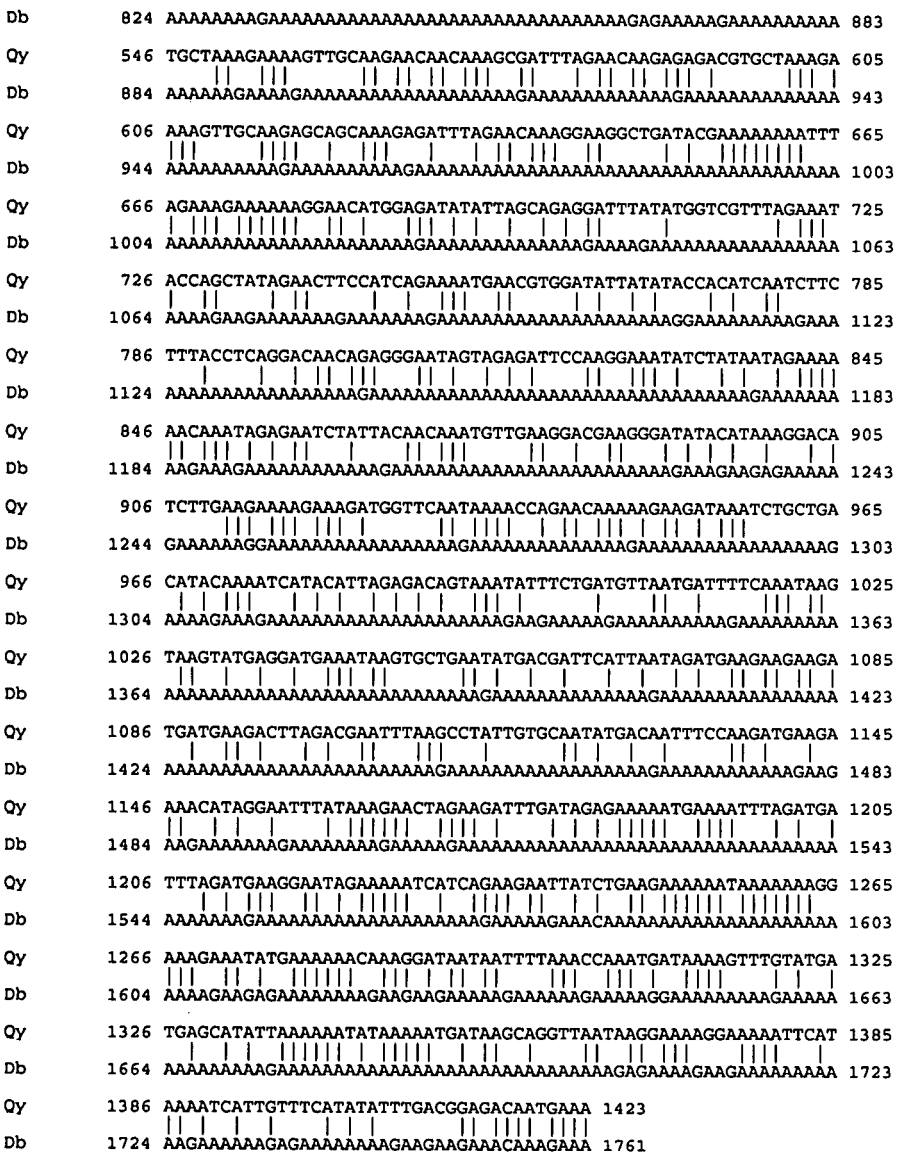




RESULT 10
DR141924
LOCUS DR141924 1783 bp mRNA linear EST 16-JUN-2005
DEFINITION 49190649 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O7 3', mRNA sequence.
ACCESSION DR141924
VERSION DR141924.1 GI:67887014
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1783)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226707030
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1783
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O7"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

Query Match 16.1%; Score 238.4; DB 8; Length 1783;
Best Local Similarity 48.0%; Pred. No. 2e-29;
Matches 680; Conservative 0; Mismatches 738; Indels 0; Gaps 0;

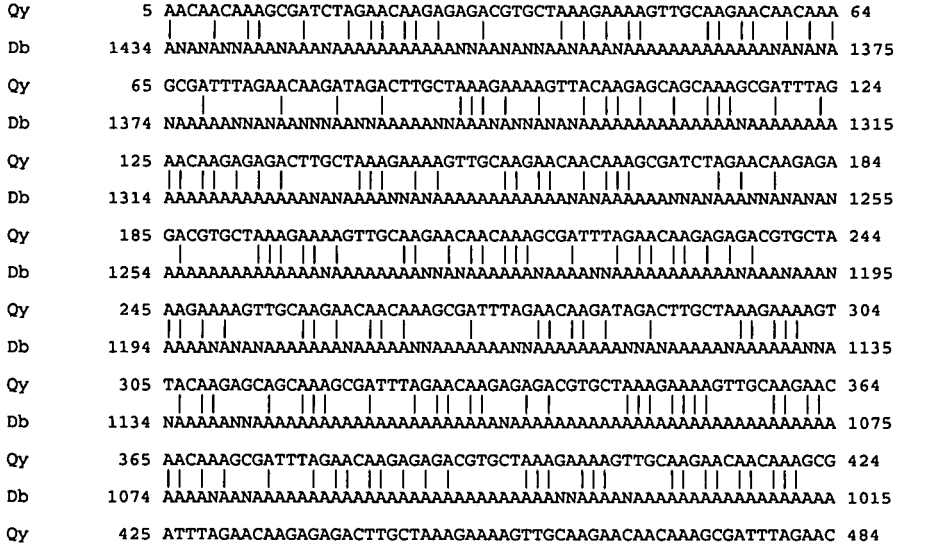


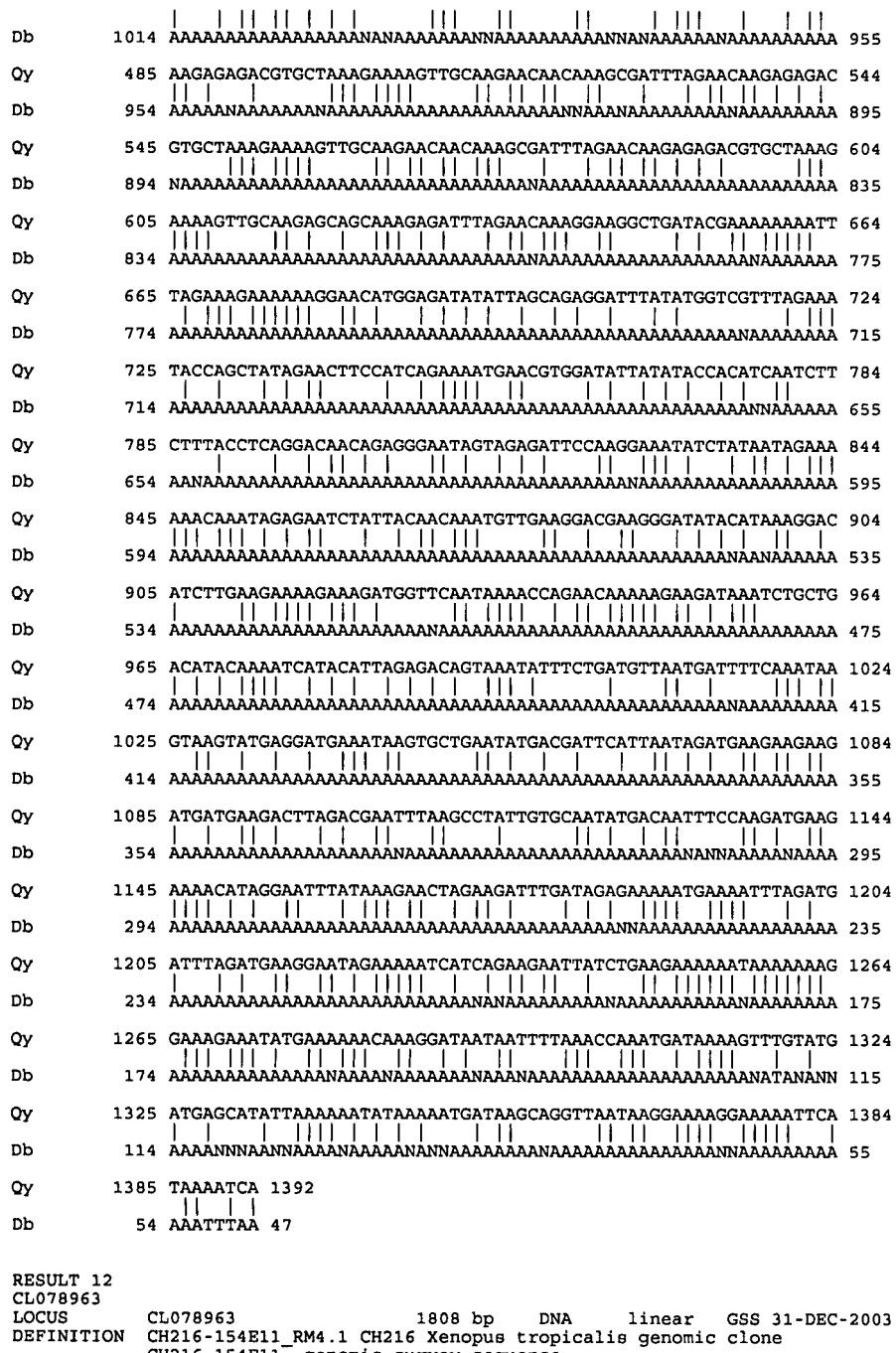


```
RESULT 11
AJ592058/c
LOCUS       AJ592058               1434 bp    DNA     linear   GSS 15-JAN-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            602D05, genomic survey sequence.
ACCESSION   AJ592058
VERSION     AJ592058.1  GI:37941682
KEYWORDS    GSS; right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 1434)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="602D05"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .1434
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN
```



```
  Query Match           16.0%;  Score 237.8;  DB 10;  Length 1434;
  Best Local Similarity   44.5%;  Pred. No. 2.5e-29;
  Matches  617;  Conservative    0;  Mismatches  771;  Indels    0;  Gaps    0;
```

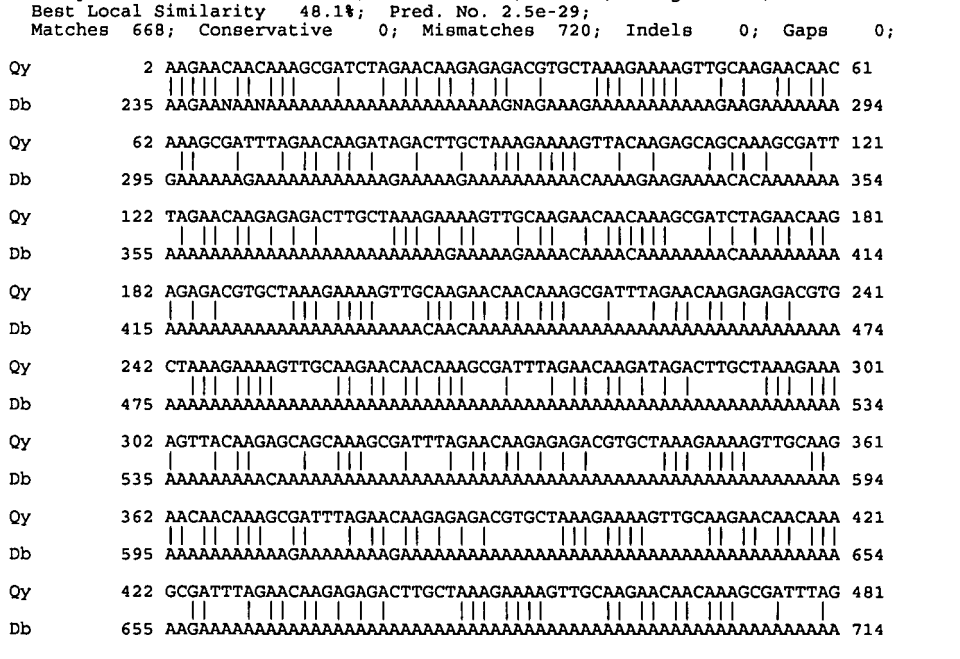


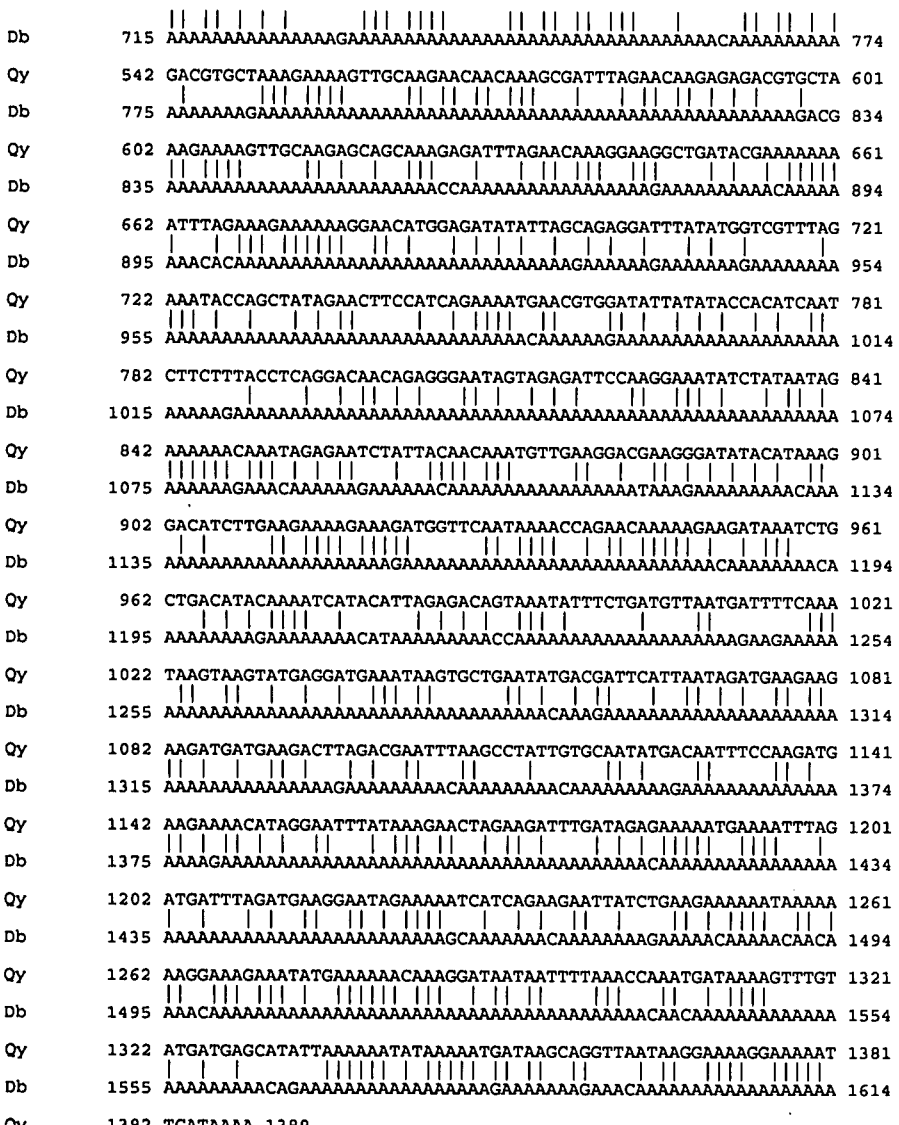


RESULT 12
CL078963
LOCUS CL078963 1808 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-154E11_RM4.1 CH216 Xenopus tropicalis genomic clone
CH216-154E11, genomic survey sequence.

ACCESSION CL078963
VERSION CL078963.1 GI:40534876
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggcatcggtcgagc
Class: BAC ends
High quality sequence start: 408
High quality sequence stop: 936.
FEATURES Location/Qualifiers
source 1. .1808
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-154E11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

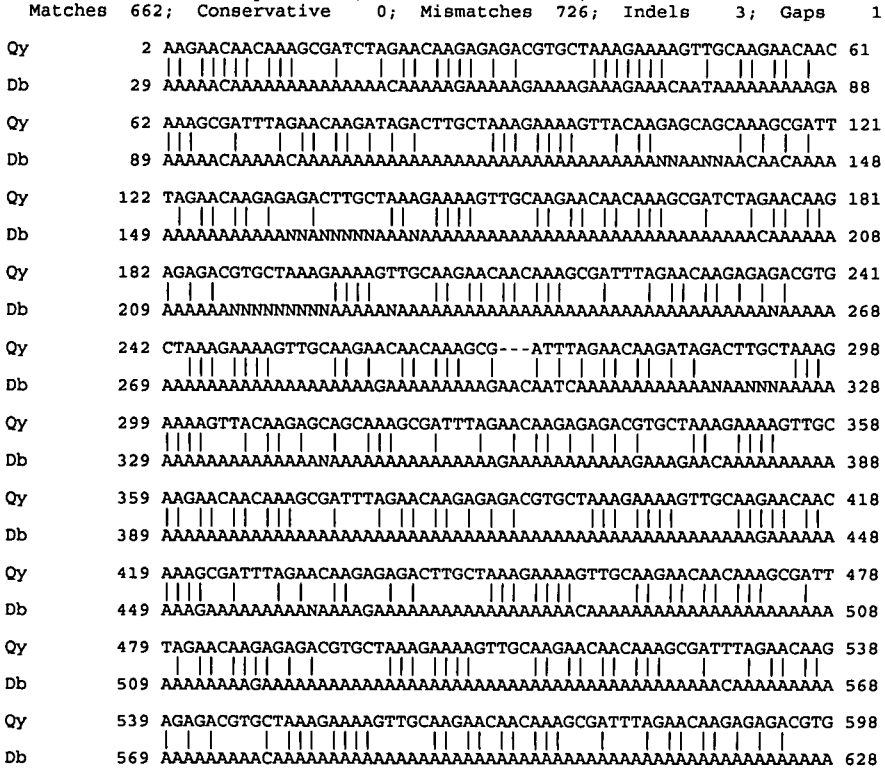
Query Match 16.0%; Score 237.8; DB 10; Length 1808;
Best Local Similarity 48.1%; Pred. No. 2.5e-29;
Matches 668; Conservative 0; Mismatches 720; Indels 0; Gaps 0;

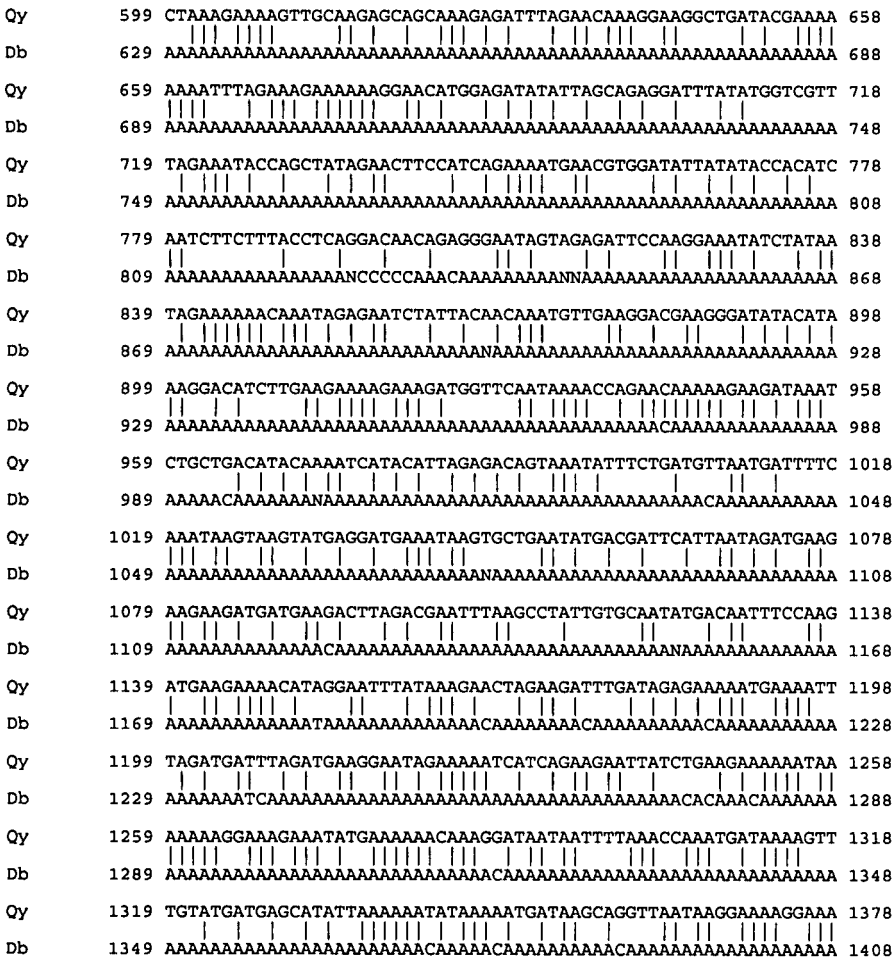




RESULT 13
CL118709
LOCUS CL118709 1654 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-72I15_T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72I15, genomic survey sequence.
ACCESSION CL118709
VERSION CL118709.1 GI:40612344
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1654)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 468
High quality sequence stop: 812.
FEATURES Location/Qualifiers
source 1. .1654
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-72I15"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

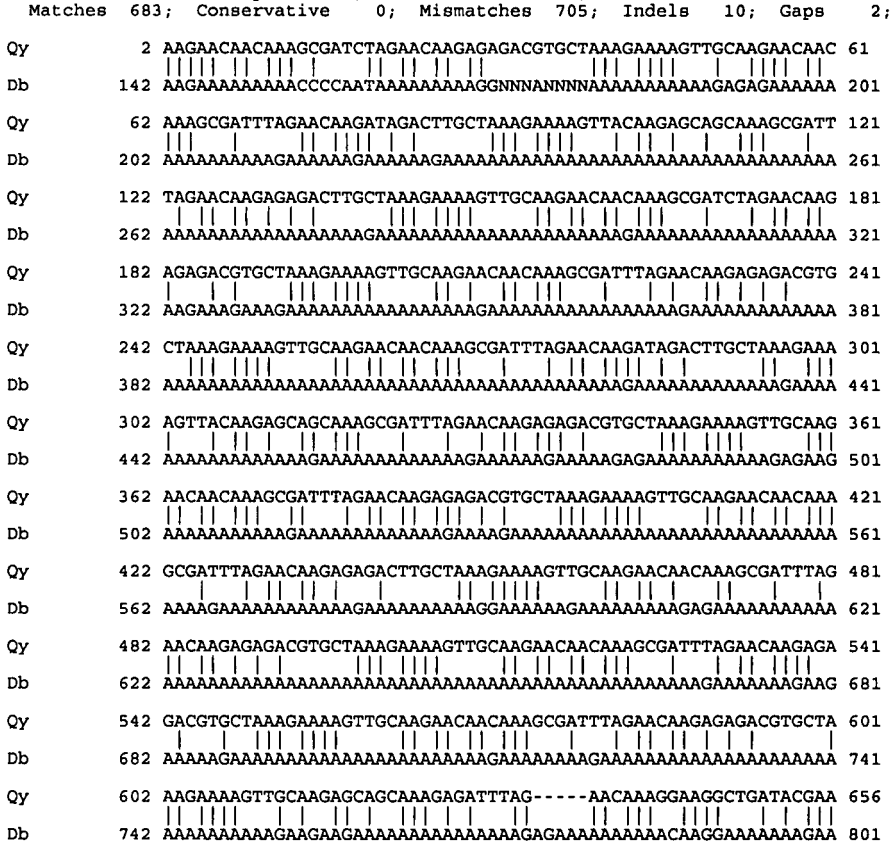
Query Match 15.9%; Score 235; DB 10; Length 1654;
Best Local Similarity 47.6%; Pred. No. 7.3e-29;
Matches 662; Conservative 0; Mismatches 726; Indels 3; Gaps 1;

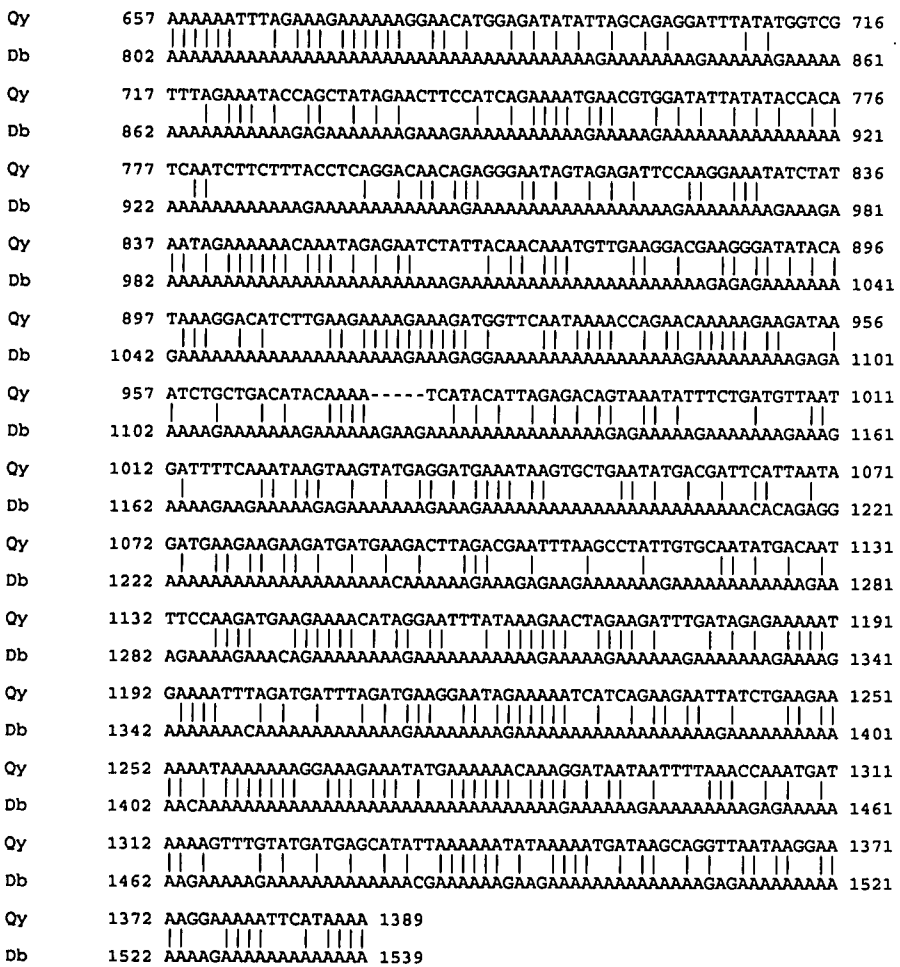




RESULT 14
CL066150
LOCUS CL066150 1797 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-108B7_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-108B7, genomic survey sequence.
ACCESSION CL066150
VERSION CL066150.1 GI:40522063
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 306
High quality sequence stop: 453.
FEATURES Location/Qualifiers
source 1. .1797
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-108B7"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

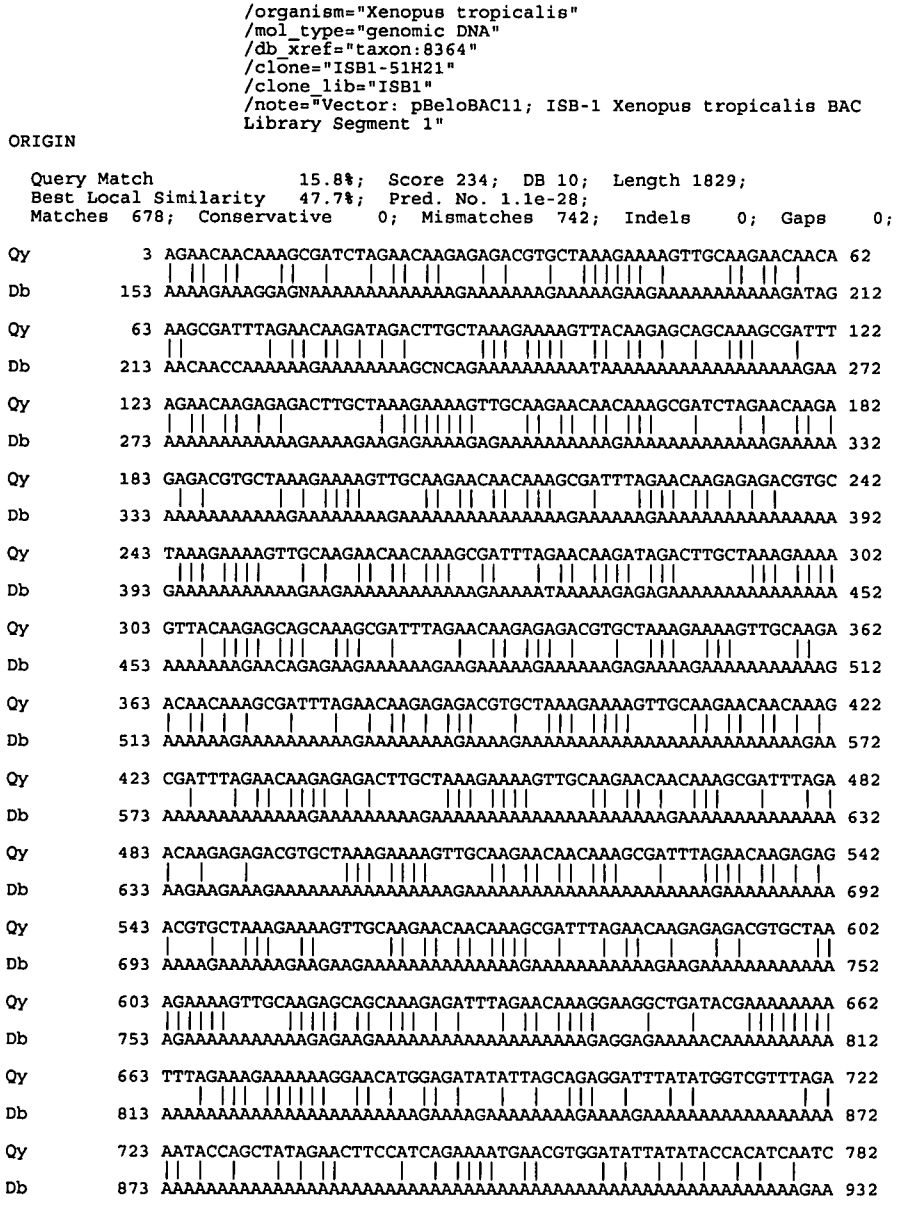
Query Match 15.8%; Score 234.2; DB 10; Length 1797;
Best Local Similarity 48.9%; Pred. No. 9.8e-29;
Matches 683; Conservative 0; Mismatches 705; Indels 10; Gaps 2;





RESULT 15
CL109409
LOCUS CL109409 1829 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-51H21_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-51H21, genomic survey sequence.
ACCESSION CL109409
VERSION CL109409.1 GI:40603044
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1829)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 856
High quality sequence stop: 936.
FEATURES Location/Qualifiers
source 1. .1829
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-51H21"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 15.8%; Score 234; DB 10; Length 1829;
Best Local Similarity 47.7%; Pred. No. 1.1e-28;
Matches 678; Conservative 0; Mismatches 742; Indels 0; Gaps 0;

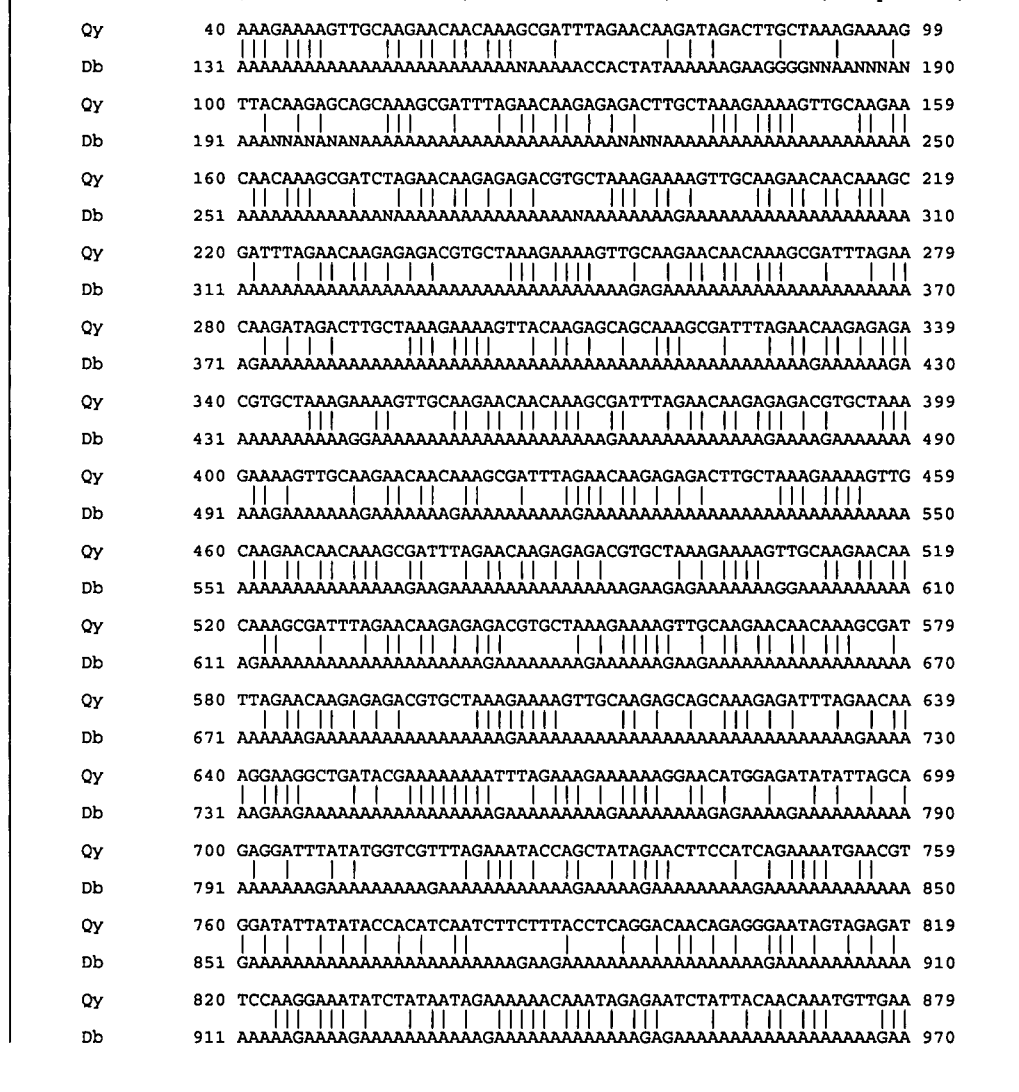


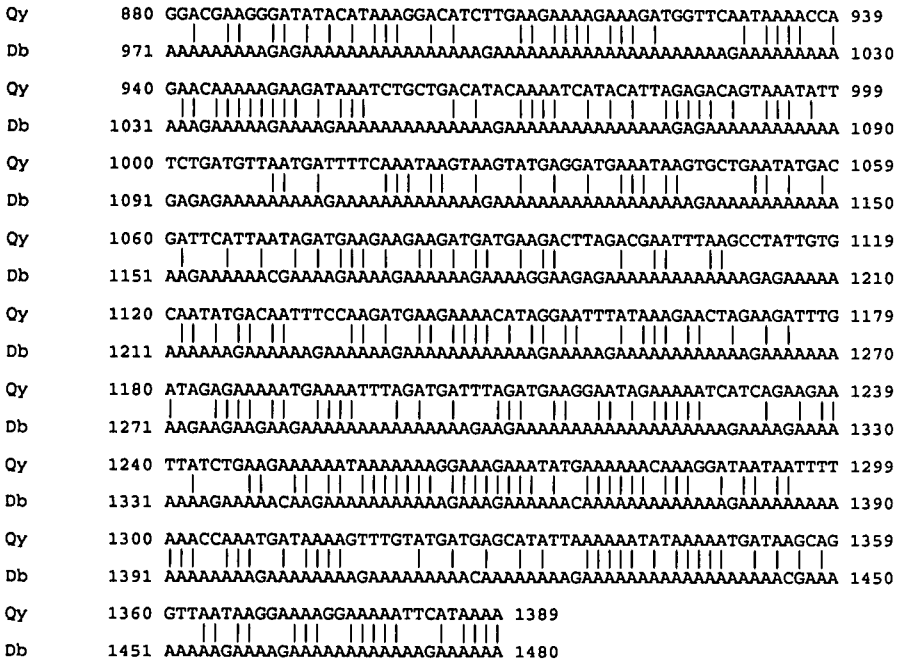


RESULT 16
CL066373
LOCUS CL066373 1736 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-108M3_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-108M3, genomic survey sequence.
ACCESSION CL066373
VERSION CL066373.1 GI:40522286
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1736)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 525
High quality sequence stop: 583.

FEATURES Location/Qualifiers
source 1. .1736
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-108M3"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 15.8%; Score 233.8; DB 10; Length 1736;
Best Local Similarity 47.9%; Pred. No. 1.1e-28;
Matches 646; Conservative 0; Mismatches 704; Indels 0; Gaps 0;



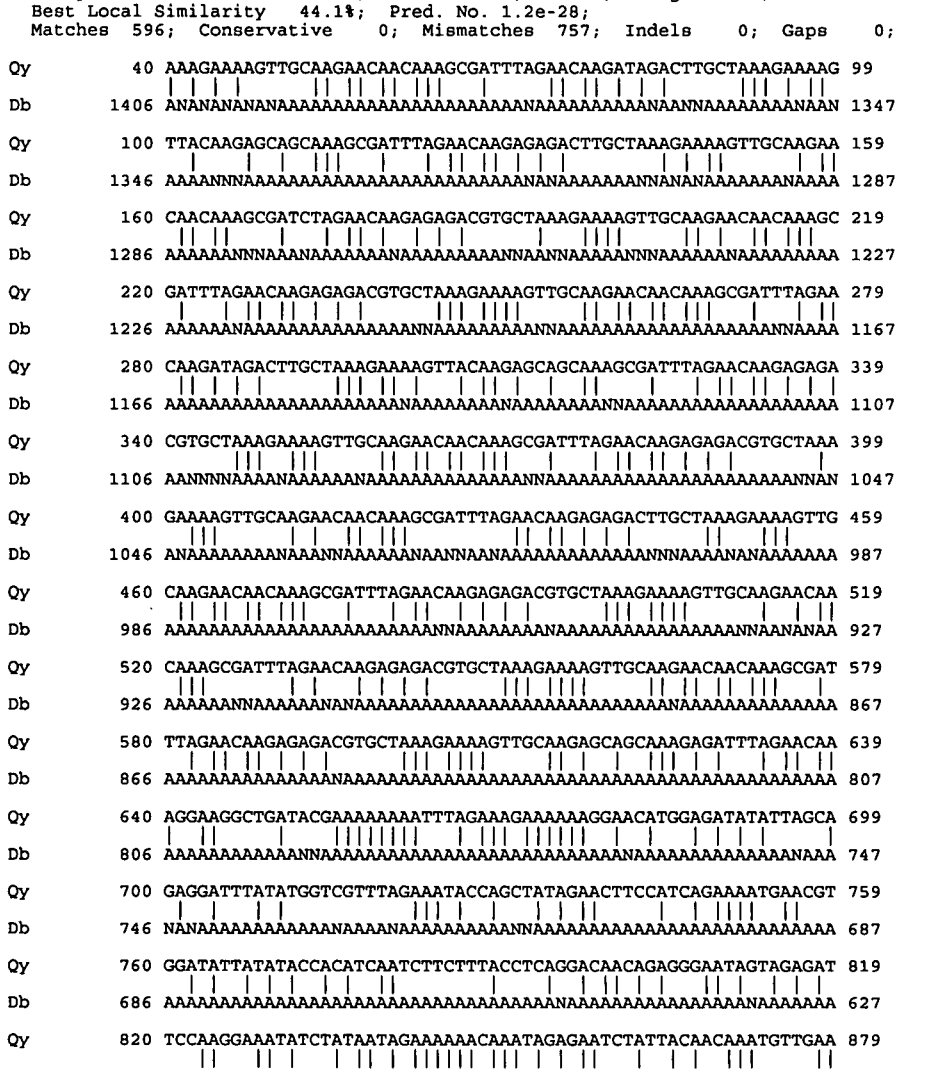


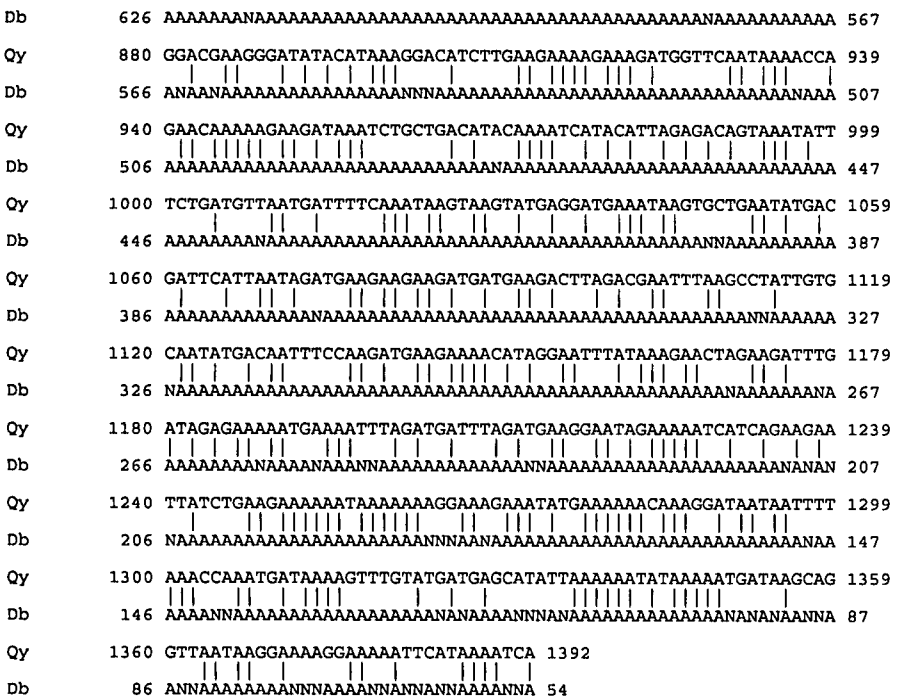
RESULT 17
AJ592026/c
LOCUS AJ592026 1407 bp DNA linear GSS 15-JAN-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 600D05, genomic survey sequence.
ACCESSION AJ592026
VERSION AJ592026.1 GI:37941650
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1407)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).

FEATURES Location/Qualifiers
source 1. .1407
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="600D05"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Wassilewskija"
misc_feature 1. .1407
/note="T-DNA flanking sequence right border"

ORIGIN

Query Match 15.8%; Score 233.6; DB 10; Length 1407;
Best Local Similarity 44.1%; Pred. No. 1.2e-28;
Matches 596; Conservative 0; Mismatches 757; Indels 0; Gaps 0;



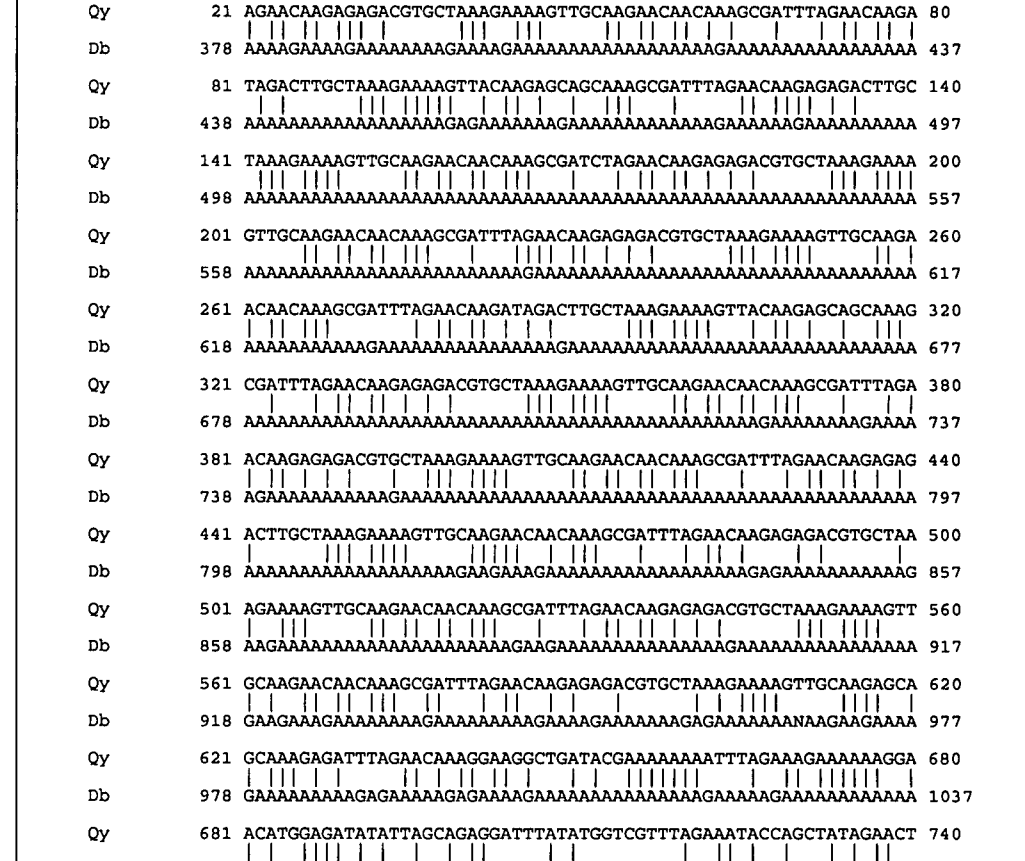


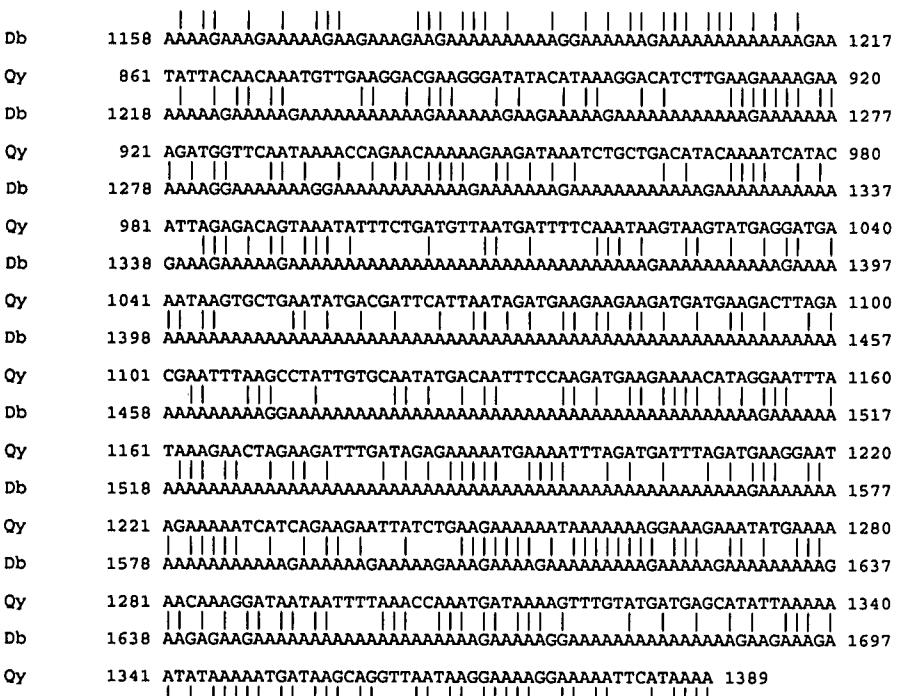
RESULT 18
DR131208
LOCUS DR131208 1753 bp mRNA linear EST 16-JUN-2005
DEFINITION 49217594 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone E9 5', mRNA sequence.
ACCESSION DR131208
VERSION DR131208.1 GI:67876120
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1753)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226709361
Insert Length: 1750 Std Error: 0.25.

FEATURES Location/Qualifiers
source 1. .1753
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="E9"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"

ORIGIN

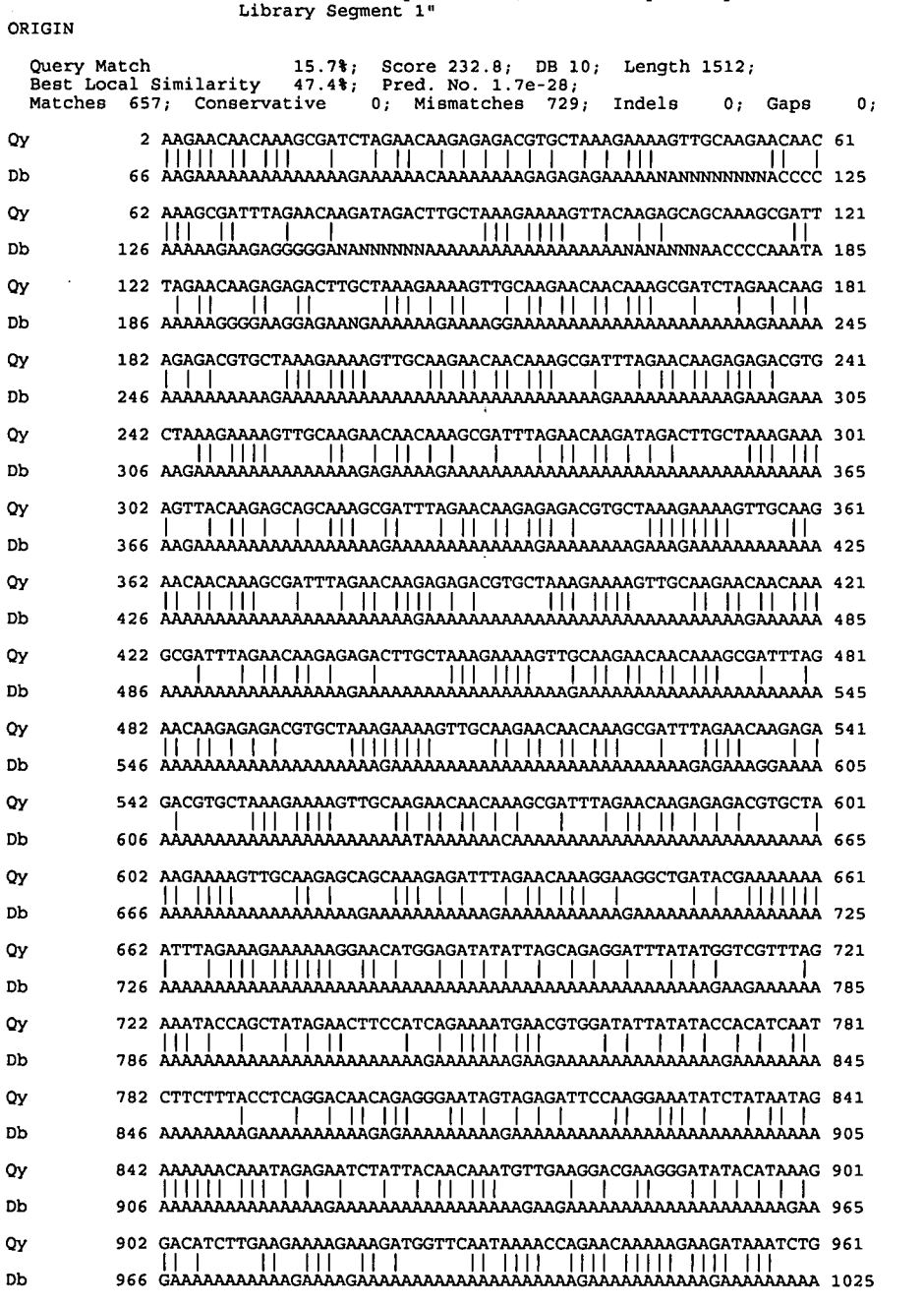
Query Match 15.8%; Score 233.6; DB 8; Length 1753;
Best Local Similarity 48.1%; Pred. No. 1.2e-28;
Matches 659; Conservative 0; Mismatches 710; Indels 0; Gaps 0;

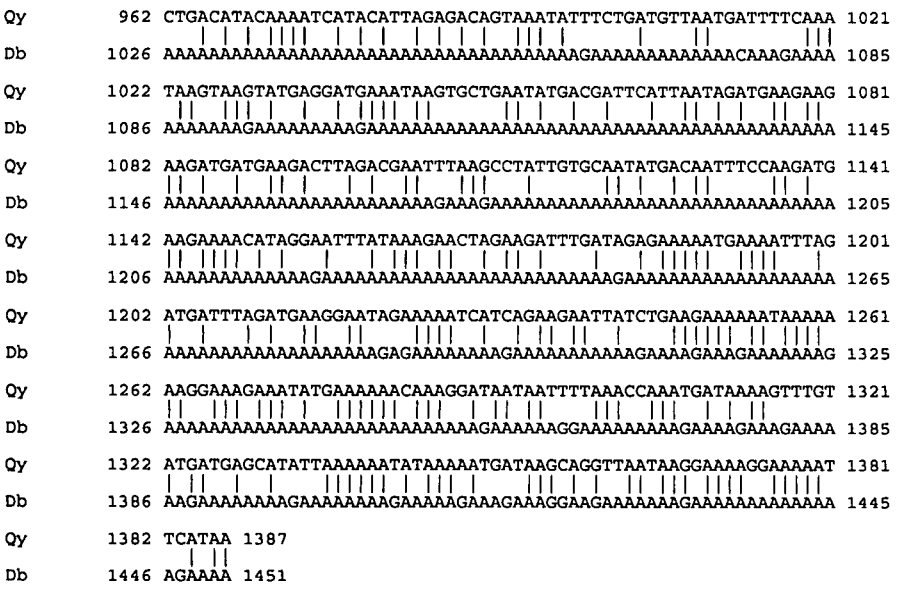




RESULT 19
CL113943
LOCUS CL113943 1512 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-59E24_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59E24, genomic survey sequence.
ACCESSION CL113943
VERSION CL113943.1 GI:40607578
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 880
High quality sequence stop: 1356.
FEATURES Location/Qualifiers
source 1. .1512
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-59E24"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 15.7%; Score 232.8; DB 10; Length 1512;
Best Local Similarity 47.4%; Pred. No. 1.7e-28;
Matches 657; Conservative 0; Mismatches 729; Indels 0; Gaps 0;

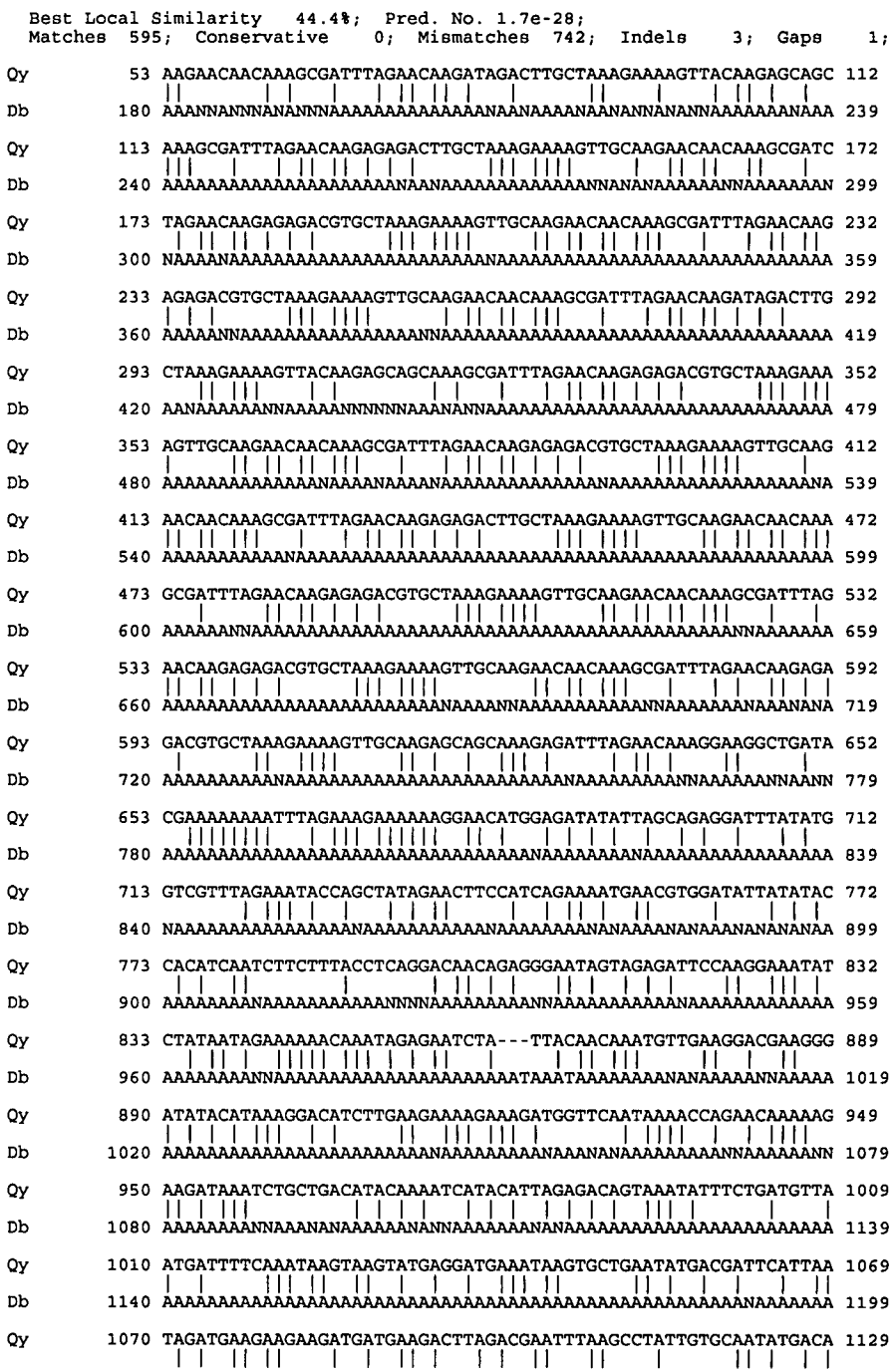


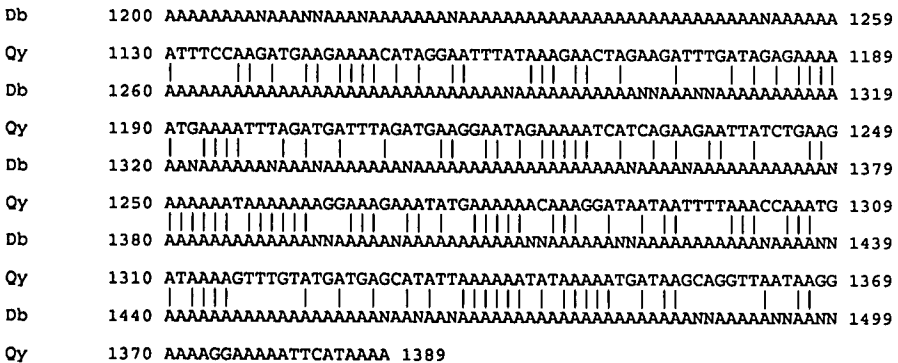


RESULT 20
CG748014
LOCUS CG748014 1531 bp DNA linear GSS 24-OCT-2003
DEFINITION P041-4-B02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG748014
VERSION CG748014.1 GI:37968940
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1531)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1531
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 15.7%; Score 232.8; DB 10; Length 1531;

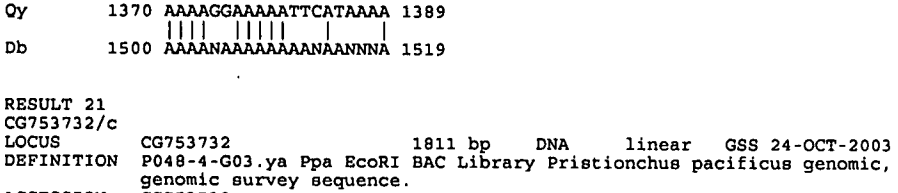
Best Local Similarity 44.4%; Pred. No. 1.7e-28;
Matches 595; Conservative 0; Mismatches 742; Indels 3; Gaps 1;

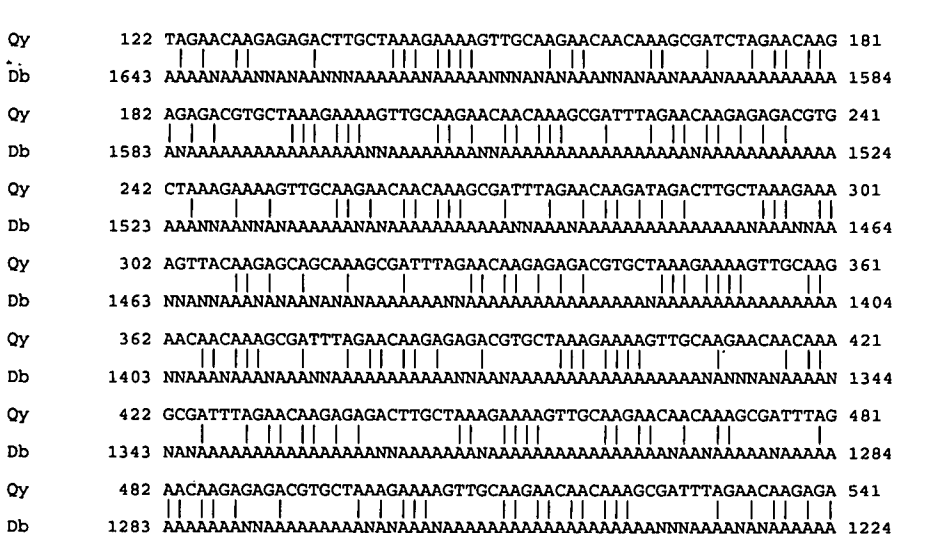


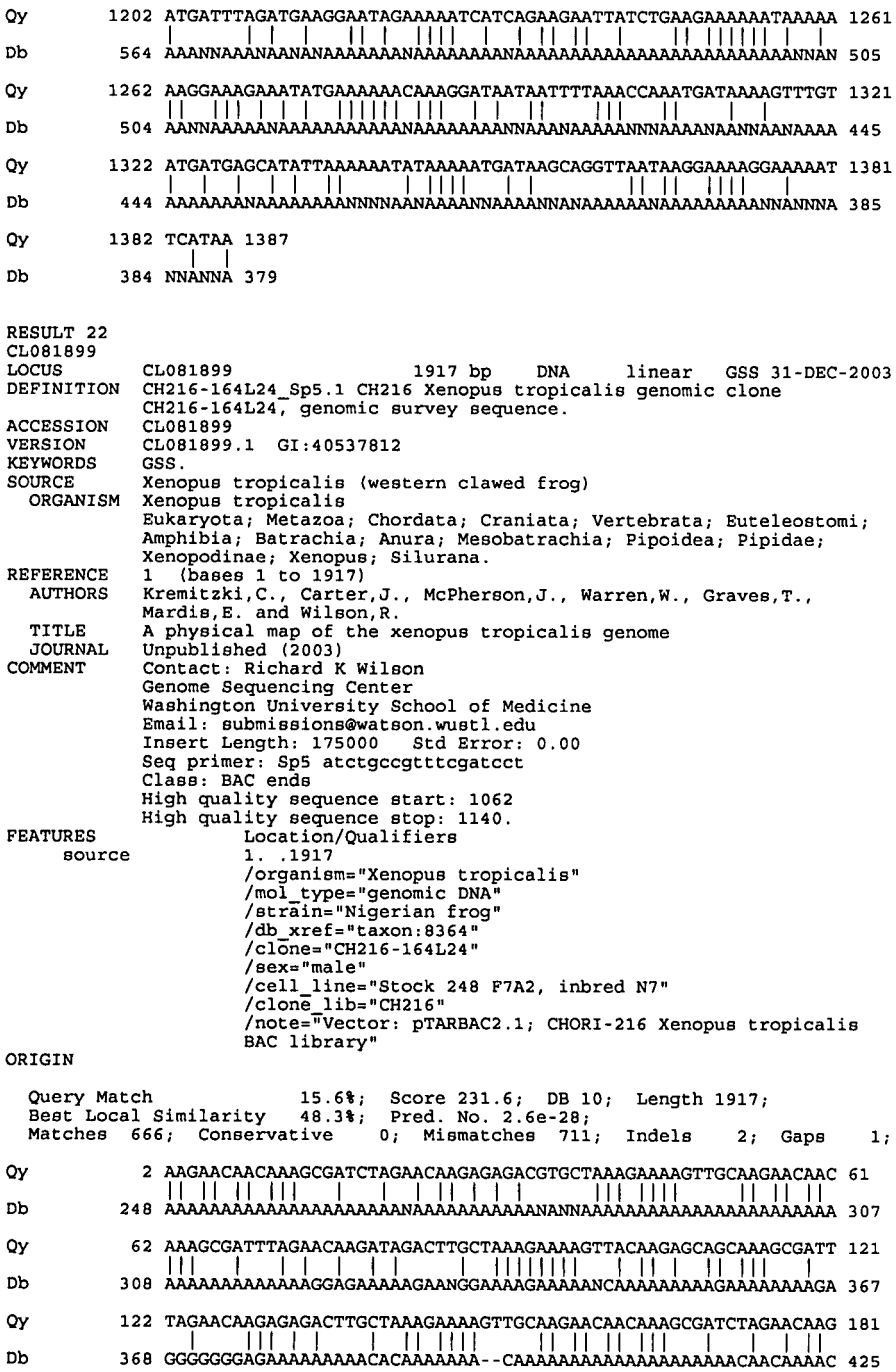


RESULT 21
CG753732/c
LOCUS CG753732 1811 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-4-G03.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753732
VERSION CG753732.1 GI:37978509
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1811)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1811
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 15.7%; Score 232.6; DB 10; Length 1811;
Best Local Similarity 40.9%; Pred. No. 1.8e-28;
Matches 567; Conservative 0; Mismatches 818; Indels 1; Gaps 1;







```
Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db        564 AAANNAAANAANANAAAAAAANAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAANNAN 505

Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
Db        504 AANNAAAAANAAAAAAAAAAANAAAAAAAANNAANAAAAANNAAAANAANNAANAAAA 445

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAAT 1381
Db        444 AAAAAAANAAAAAAAANNNNAANAAAANNAAAANNANAAAAAANAAAAAAAAANNANNNA 385

Qy       1382 TCATAA 1387
Db        384 NNANNA 379


RESULT 22
CL081899
LOCUS       CL081899               1917 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-164L24_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-164L24, genomic survey sequence.
ACCESSION   CL081899
VERSION     CL081899.1  GI:40537812
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1917)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1140.
FEATURES             Location/Qualifiers
     source          1. .1917
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-164L24"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.6%;  Score 231.6;  DB 10;  Length 1917;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-28;
  Matches  666;  Conservative    0;  Mismatches  711;  Indels    2;  Gaps    1;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db        248 AAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAANNAAAAAAAAAAAAAAAAAAA 307

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db        308 AAAAAAAAAAAAAGGAGAAAAAGAANGGAAAAGAAAAANCAAAAAAAAAGAAAAAAAAGA 367

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db        368 GGGGGGGAGAAAAAAAAACACAAAAAAA--CAAAAAAAAAAAAAAAAAAAACAACAAAAC 425

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db        426 ACAAACAAAAAAACAAAAAAAAAAAAAACAAAACAAAGAAAAGANAAAAAAGAAAAAAGAA 485

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db        486 GAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAAAAAGAAAAA 545

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db        546 AAAAAAAGAAAAAGAAAAAAAAAAAAAAAGAAAAAAAAAAAAAACAAAAAAAAAAAGAAAAA 605

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db        606 AAAAAAAAAAAAAAAAAGAAAAAGAAAAAAGAAAAAAAAAAGGAGAAAAAAAAAAAAAGAAA 665

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db        666 AAAAAAAAAAAGAAAAAAAAGAAAAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 725

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db        726 AAGAAAAAAGAAAAAAAAGAGAAAAAGAAAAAGGAGAAAAAAAAAAAAACAAAAAAAAAAG 785

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db        786 AAAAAAAAAAAGAAAAAGAAAAAAAAAAAACAGAAAAAAAAAAAAAAAGAGAAAAAAAAAAAA 845

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db        846 AAGACAGAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAGAGAAAAAAAGAAAGAAAA 905

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db        906 AGAAAAAAAAAAGAAAGAATAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGAAAAAA 965

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db        966 AAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAGAAAAAA 1025

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAG 841
Db       1026 AAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAGAAAAAAA 1085

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db       1086 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAA 1145

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db       1146 AAAAATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAGAAAAAAAAAAAAAAAAAAG 1205

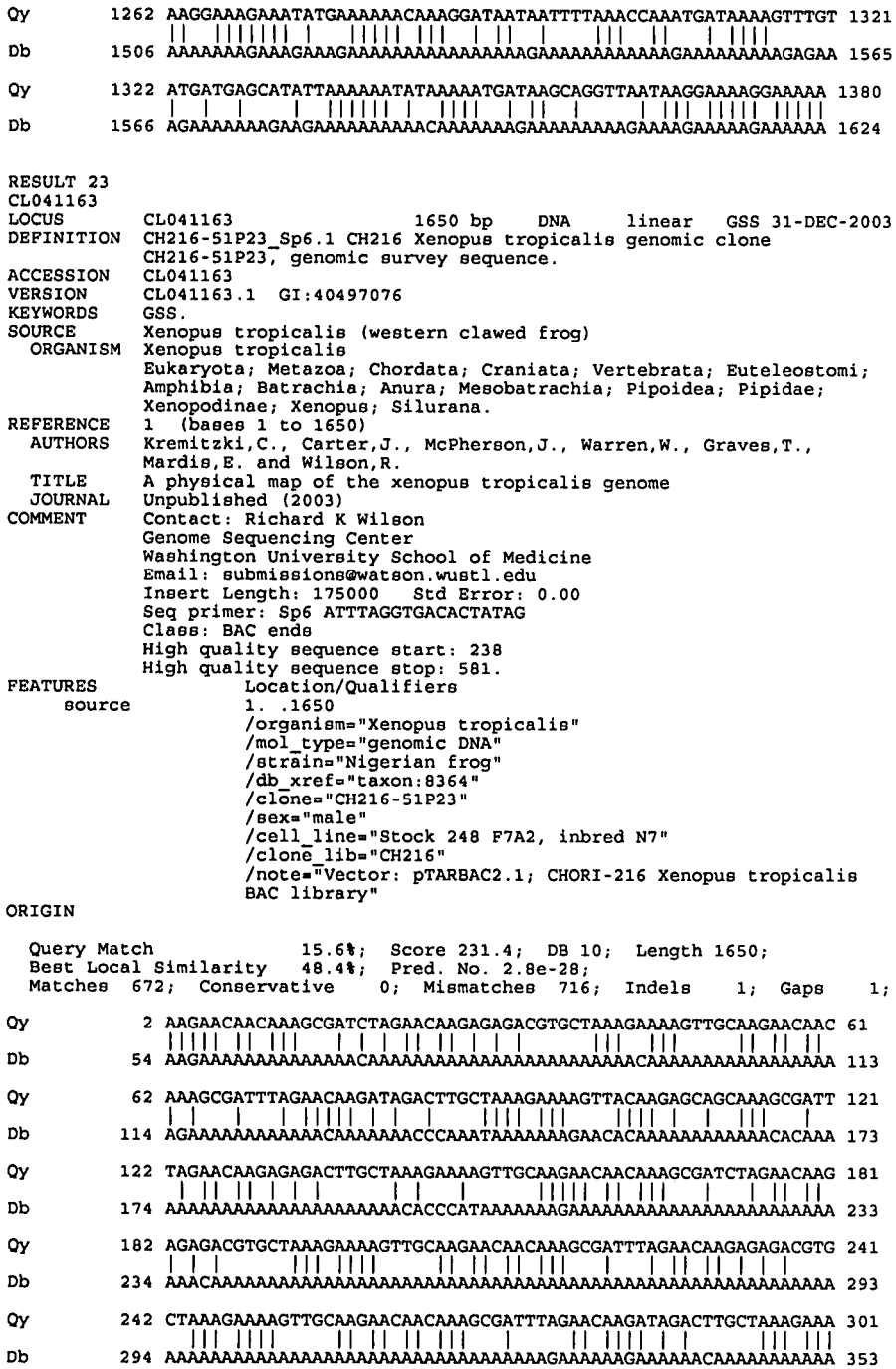
Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       1206 ACAAAGAAAAAAAAAAGAAAGAAAAAAAAGAAAAAAAGAAAAAAGAAAAAAAAAAAAAAA 1265

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db       1266 AAAAAAAGAAAGAAAAGAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAAAAAAGGAAAA 1325

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
Db       1326 AAGAAAAAGAAAAAAAAAGCAAAGAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAA 1385

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
Db       1386 AAAAAAAAAAAAGAAAAAAAAAGAAAGAAAAAAAAAAAAAAAAAGAAGAAAAAAAAAGAAG 1445

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db       1446 AAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAGAAACAAAAAAAGAAAAAAAAAAAAAA 1505
```



RESULT 23
CL041163
LOCUS CL041163 1650 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-51P23_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-51P23, genomic survey sequence.
ACCESSION CL041163
VERSION CL041163.1 GI:40497076
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1650)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 238
High quality sequence stop: 581.
FEATURES Location/Qualifiers
source 1. .1650
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-51P23"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 15.6%; Score 231.4; DB 10; Length 1650;
Best Local Similarity 48.4%; Pred. No. 2.8e-28;
Matches 672; Conservative 0; Mismatches 716; Indels 1; Gaps 1;

```
Qy       1381 TTCATAAAA 1389
              | ||||
Db       1434 AAAAAAAAA 1442


RESULT 24
CL082587
LOCUS       CL082587               1762 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-168C5_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-168C5, genomic survey sequence.
ACCESSION   CL082587
VERSION     CL082587.1  GI:40538500
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1762)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 1020
            High quality sequence stop: 1162.
FEATURES             Location/Qualifiers
     source          1. .1762
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-168C5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.6%;  Score 231.2;  DB 10;  Length 1762;
  Best Local Similarity   48.0%;  Pred. No. 3.1e-28;
  Matches  713;  Conservative    0;  Mismatches  766;  Indels    7;  Gaps    2;

Qy           2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
               || || || |||   |   | || ||   | |   | | || ||      || || ||
Db         269 AAAAAAAAAAAAAAAAANAAAAAAAAANNNAAAAAAGAGAGAGGAAGAAAAAAAAAAAAA 328

Qy          62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
               |||  ||   | || || | | |       | |  |||   | || |  |  ||| ||
Db         329 AAAAAGAGAAAAAAAAAAAAAAAAAAAAAAAAANAAGAAAAGAAAAAAAAAAAAAAAAAGAAA 388

Qy         122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
                 | || || | | |       | ||||||    ||||| || |||   |   | || ||
Db         389 AAAAAAAAAAAAAAAAAAAAGAGGAAAAGAAAAAAGAAAAAAAAAAAAAAGAGAAAAAAAA 448

Qy         182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
               | | |       ||| ||||      || || || |||  ||   | || |||| | |
Db         449 AAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAAAAGAAAAAAAAA 508

Qy         242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
                |||||  ||      || || |  |||        | || ||   | |       ||| ||
Db         509 AAAAAGAGAAAAAAAAAAAAAAAGAAAAAAAGAAAAAAAGAAAGAAAAAAAAANGAAAAAAG 568

Qy         302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
               |   | || |  || |||   |  | || || | | |       ||| ||||      |||
Db         569 AAAAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 628

Qy         362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
               || ||  ||| ||   | || || | |||        ||||||      | ||  || ||
Db         629 AAAAAAGAAGAGAAAAAAAAAAAAAAAAAGAAAAANAGGAGAAAAAAAAAGAAAAACACCAA 688

Qy         422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
                 |  | ||  | |||| |     ||  ||||   | || || || |||   |   |
Db         689 AAAAAAGAAAAAGAAAGAGGCAAAAAAAGAAAAAAAGGAAAAAAAAAAAAAAAACAAAAAA 748

Qy         482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
               ||     |           ||| |||       || ||  | | |  ||   | || || | |
Db         749 AAAGGAAAGAGAAAAAAAAAAAAAGAAAAAAAAAAAGAAAGAAAGAAAAAAAAGAAAAAA 808

Qy         542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
                 |     ||| |||||  | |  ||  | |||    |   |||| || | |||
Db         809 AAAAAAAGAAAAAAAAGAAGAAGAAAAGAAAAAAAAAAGAAAGAAAAAAAAAAGAAAAAAAG 868

Qy         602 AAGAAAAGTTGC------AAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
               || |||||          || || |  ||| | |   | || |||| ||| |  | |  |
Db         869 AAAAAAAGCAAAGAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAGAAAGACAAAAAAAA 928

Qy         656 AAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTC 715
               |||||||   | ||||||||||  || |      | | |  || | |  |      |
Db         929 AAAAAAAAGAAAAAAGAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAGAAAAAAA 988

Qy         716 GTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCAC 775
                 | ||| |  ||    | ||     |  | ||||  ||     | |  | | |  |
Db         989 AGAGAAAAAGAAAAGAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAAAA 1048

Qy         776 ATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTA 835
               |  ||         |    || | || |||   || |  | | |    ||  ||| |   |
Db        1049 AAAAAGAAAAAAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1108

Qy         836 TAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATAC 895
               || | |||||  || | ||  |  | || |||      ||  |  ||  | | |
Db        1109 AAAAAAAAAAAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAA 1168

Qy         896 ATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATA 955
               | ||| || |    || |||| ||| |  |   || ||||  | || |||||    ||
Db        1169 AAAAAAGAGAAGAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAACGGAAG 1228

Qy         956 AATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATT 1015
               ||       | | ||||  |   |  | | | |  |||       || |  ||  |
Db        1229 AAAAGAAGAAGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAAAGAAGAAAAAAAAG 1288

Qy        1016 TTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATG 1075
                  ||| ||  ||  |   | |  ||| ||      || |  |  |   | || | |
Db        1289 AAAAAAAAAAGAAAAAGAGGAAAAAAAAAAAAAAAGAGAAAAAAAAAAAAAAAAAGAAAAA 1348

Qy        1076 AAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCC 1135
               || || ||||  |  || |  |  |  ||  ||     |      || |  | ||
Db        1349 AAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAAAAAAACAGA 1408

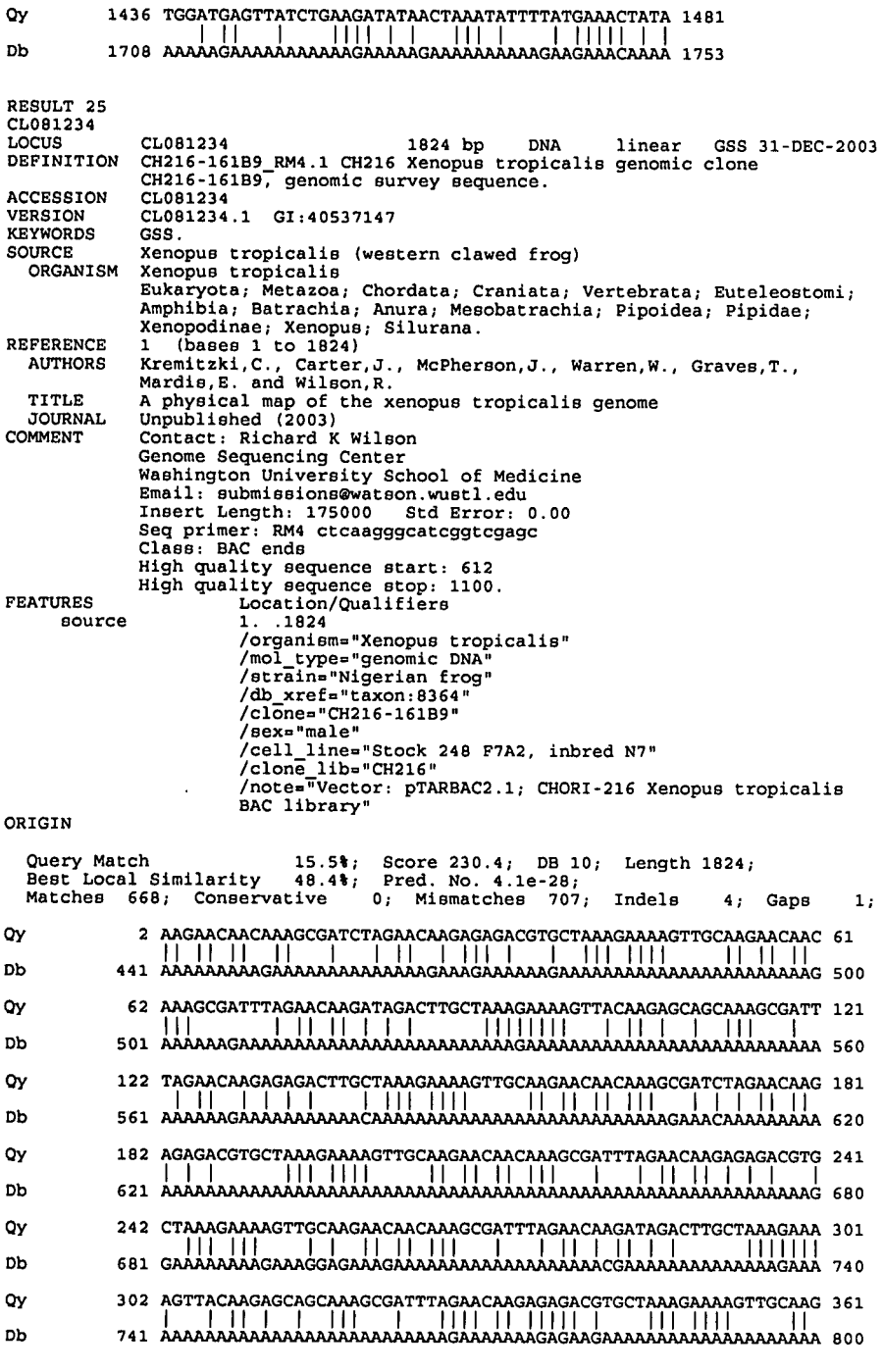
Qy        1136 AAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAA 1195
               || |  |||||| | |  ||   |  || ||  |||| |    | | | |||||  |||
Db        1409 AAAAGAAAGAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAA 1468

Qy        1196 ATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAA 1255
               |  | | ||   | |  || ||| |  | ||  |  | ||||  |    || ||||||
Db        1469 ACGAAAAAGAAAAAAAAAAAAGAA-AAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAA 1527

Qy        1256 TAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAA 1315
               |||  |||  |||||||  |  || ||| | | || || ||     | |   ||   ||||
Db        1528 AAAACAAAAAAAAGAAAGAAAAAGAAAAAGAAGAGAAAAAAAAAAAGAAAGAAAAGAAAAA 1587

Qy        1316 GTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGG 1375
               |   | |  |   |  |     |||||| | | |||  | ||  |     || |   |||||
Db        1588 GAAAGAAAAAAAAAAAAACAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAAAAAGAAAAAAGA 1647

Qy        1376 AAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCG 1435
               |||||   | ||||  |       | | |    | |||| || |||    | |
Db        1648 AAAAAAAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAGAGAGAAAGAAGAAAGAAAAGAAA 1707
```



```
Qy       1436 TGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATA 1481
Db       1708 AAAAAGAAAAAAAAAAAGAAAAAGAAAAAAAAAAAGAAGAAACAAAA 1753

RESULT 25
CL081234
LOCUS       CL081234                1824 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-161B9_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-161B9, genomic survey sequence.
ACCESSION   CL081234
VERSION     CL081234.1  GI:40537147
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1824)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 612
            High quality sequence stop: 1100.
FEATURES             Location/Qualifiers
     source          1. .1824
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-161B9"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.5%;  Score 230.4;  DB 10;  Length 1824;
  Best Local Similarity   48.4%;  Pred. No. 4.1e-28;
  Matches  668;  Conservative    0;  Mismatches  707;  Indels    4;  Gaps    1;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db        441 AAAAAAAAAGAAAAAAAAAAAAAAGAAAGAAAAAAGAAAAAAAAAAAAAAAAAAAAAAG 500

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db        501 AAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAA 560

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db        561 AAAAAAGAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAACAAAAAAAA 620

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db        621 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 680

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db        681 GAAAAAAAAGAAAGGAGAAAGAAAAAAAAAAAAAAAAAAACGAAAAAAAAAAAAAAAGAAA 740

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db        741 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAGAGAAGAAAAAAAAAAAAAAAAAA 800

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db        801 AAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAA 860

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db        861 AAAAAAAACAAAAAAAAAAAAAAAAAAAAAAGAAAAAGAGAAAAAAAAAAAAAAAAAAAAA 920

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db        921 AAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAACA 980

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db        981 AAAAAAAAAAAAGAAAAAAGAAAAAAAAGAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAA 1040

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db       1041 AAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAAAA 1100

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db       1101 AAGGAAAAAAAAAGAAGAAAAAAGAAAAAGAAAAAAAAAAAAAAAAAAAAAGAAAAAAAA 1160

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db       1161 AAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAA 1220

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAG 841
Db       1221 AAAGAGAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAGAA 1280

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db       1281 AAAAGAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAA 1340

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db       1341 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAA 1400

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       1401 AAAAAAAAAAAAAGGAAAAGAAAAAAAAAGAAAAAAAAAACAAGAAAAAAAAA----AAA 1456

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db       1457 AAGAAAAAAAAAAGAACAAAAGAAGAAAAAAAAAAAAAGAGATAGAAAAAAAAAAAAAAA 1516

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
Db       1517 AAAAAAAAAAAAAGAGAGAAAAAAAAAAANGAAGAAAAAAAAAAAAGAAAAGAAAAAAA 1576

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
Db       1577 AAAAAAAAAAAAAAGGAAAAGAAAGAGAAAAAGAAAAGACAGGAAAAAGAGAAAAAAAAG 1636

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db       1637 GATAAAAAAAAAAACAAAAAAGAGAGAAAAAAAAGAAAAAAAAGAAAAAAAAAAAGAAAA 1696

Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
Db       1697 AAAAAAAAAAAAAAAAAAAAAGAAGACAAAAACAAAAAAAAGGAAAAGAGAAAAAAAACA 1756

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1757 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAAGAAAAAAAAAAAAAAA 1815

RESULT 26
CL118815
LOCUS       CL118815                1727 bp    DNA     linear   GSS 05-JAN-2004
```

```
DEFINITION  ISB1-72O5_T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72O5,
            genomic survey sequence.
ACCESSION   CL118815
VERSION     CL118815.1  GI:40612450
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1727)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 903
            High quality sequence stop: 1314.
FEATURES             Location/Qualifiers
     source          1. .1727
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-72O5"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             15.5%;  Score 230.2;  DB 10;  Length 1727;
  Best Local Similarity   47.9%;  Pred. No. 4.5e-28;
  Matches  664;  Conservative    0;  Mismatches  723;  Indels    0;  Gaps    0;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              | || || ||  | ||  | || |  | | |   | ||| ||||     || || ||  |
Db        272 AAAAAAAAAACACAATAAAAAAAACAAAAAAAAAGGAAAAAAAAAAAAAAAAAAAAAAA 331

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              |  | ||  | || | || | |       ||| ||||   | || |  |  |||
Db        332 ACACAATAAAAAAAATGAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAGAAA 391

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
                || || | | |    | ||| ||||     || || || |||   |   | || || |
Db        392 GAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAACAAAAAAAAAAACACAAAAAAAAAA 451

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              ||         ||| ||||     || || || |||   |   | || || | | |  |
Db        452 GAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA 511

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
               ||| ||||    | || |  ||||||    |      || || | | |      ||  ||||
Db        512 AAAAAAAAAAAAGAAAATAAACAAAAAAAAAATAAAAAACACAAAAAAAAGAACAAAAA 571

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
                | || |  |  |||   |   | || || | | |      ||| ||||     || |
Db        572 AAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAACAAACAAAAAAAAAAAAAAAA 631

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | || |||   |   | || |||| | |      ||| || |     || || || |||
Db        632 AAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAACAAAAAAAAAGAAAA 691

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
                |   | | || | | |      ||| |||      ||  |  | |||  ||   | |
Db        692 AAAGAAAACAAAAAAAAAAAAAAACAAAAAAATAAAAAAAATAAGAAAAAAAGAAAAAA 751

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              |||| | | |   |  ||| ||||     || || |  |||   |   | || |||| |
```

```
Db        752 ACAAAAAAAAGAAGACAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAACAAAAAGAAAA 811

Qy        543 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 602
              |   |  ||  ||||     || || || |||   |   | ||  ||| |         ||
Db        812 AGAAGAAAAGAAAAAAAAAAAACAATAAAAAAAAAAAAAAAAAAACAGAAAACAAAAAAAA 871

Qy        603 AGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAA 662
              | ||||   | || |  | | | | |   | || |||  ||      | |  ||||||||
Db        872 AAAAAAAAAGAAAAAAAAAAACAAAAACAGAAAAAAAAAAAAAAAAAAAAAACAAAAAAAA 931

Qy        663 TTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGA 722
                | ||| ||||||  || |   | | | |  | |||  |  | |        | |
Db        932 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAA 991

Qy        723 AATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATC 782
              || |  | | | | ||     |  | ||||  ||     | |  | | |  | |  ||
Db        992 AAGAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1051

Qy        783 TTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGA 842
                  |   | | || | |   ||     | | |      |  ||| |   | || | |
Db       1052 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAACAAAAAAAAAAAAAAAAA 1111

Qy        843 AAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGG 902
              ||||| ||| | | ||   |  | || |||     ||  |  ||   | | | | ||| |
Db       1112 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 1171

Qy        903 ACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGC 962
              | |     ||  ||| |||       || ||||  | ||  ||||| ||||  ||
Db       1172 AAAAAAAAAAACAAAAAAAAGAAAGAAAAAAAAAAAAAAAGAAAAAAAAGAAGAAAAACA 1231

Qy        963 TGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAAT 1022
              || | | ||||  | | |  | | | |  ||| |       |     ||  |     |||
Db       1232 AGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1291

Qy       1023 AAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGA 1082
              ||  ||  |  |  |  ||| ||      || |  |  |   |  ||   |  || || |
Db       1292 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAA 1351

Qy       1083 AGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGA 1142
                |  |  |  |  | |  ||  ||  ||      ||    | ||     ||    |
Db       1352 GAAAAAAAGACAAAAAAAGAAAAAAAACAATAAAAAAAAAAGAAAAAAAAAAAAAATAAA 1411

Qy       1143 AGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGA 1202
              | |||| | |  ||   | ||| |  | || |       | | |||||  ||||   | |
Db       1412 AAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAACA 1471

Qy       1203 TGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAA 1262
                |  | |  ||  || | ||||   |  | || ||  |    || ||||||  || |||
Db       1472 AAAAACAAAAAAAAAAAAAAAAAAGAAAAAATAAAAAAAAAACAAAAAAAAAAAAACAAA 1531

Qy       1263 AGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTA 1322
              |   || ||| |  |||||| |||  | || ||    |||| |||  | ||||    | |
Db       1532 AAACAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAACGAAAAAAAAAAAAAAAGAA 1591

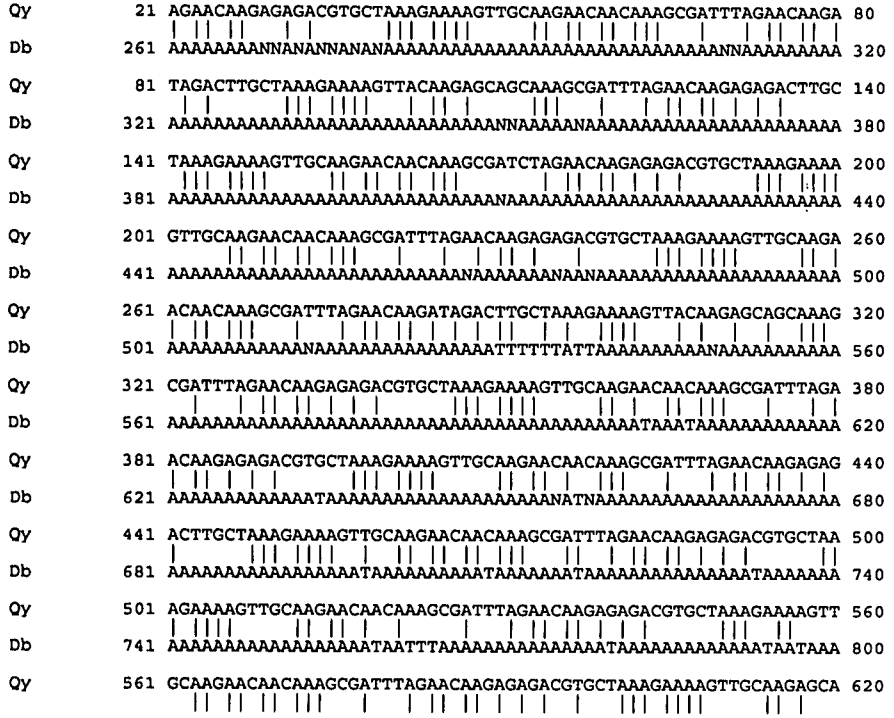
Qy       1323 TGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATT 1382
                |  |  | |  |||||| | ||||| || ||  |    ||  |  |||    | ||
Db       1592 GAAAAAAAAAAGAAAAAAAAAAAAAAAAGAAAAAAAAAACAAATAAAAAAGAACATAAAA 1651

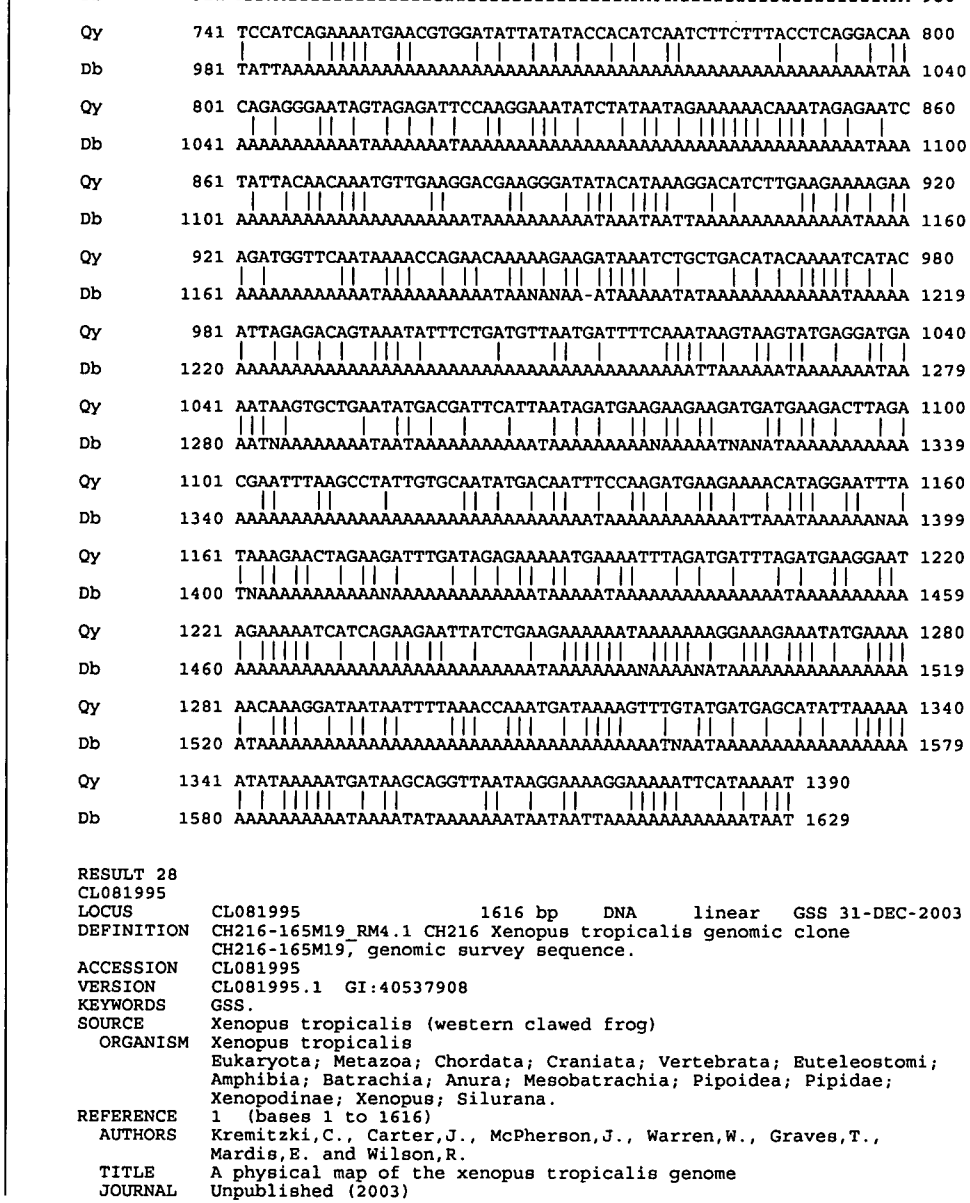
Qy       1383 CATAAAA 1389
              |  ||||
Db       1652 CGAAAAA 1658
```

```
RESULT 27
CL082569
LOCUS       CL082569                1632 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-167P5_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-167P5, genomic survey sequence.
ACCESSION   CL082569
VERSION     CL082569.1  GI:40538482
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1632)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1114.
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-167P5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match           15.5%;  Score 230;  DB 10;  Length 1632;
  Best Local Similarity   47.6%;  Pred. No. 4.8e-28;
  Matches  652;  Conservative    0;  Mismatches  717;  Indels    1;  Gaps    1;
```

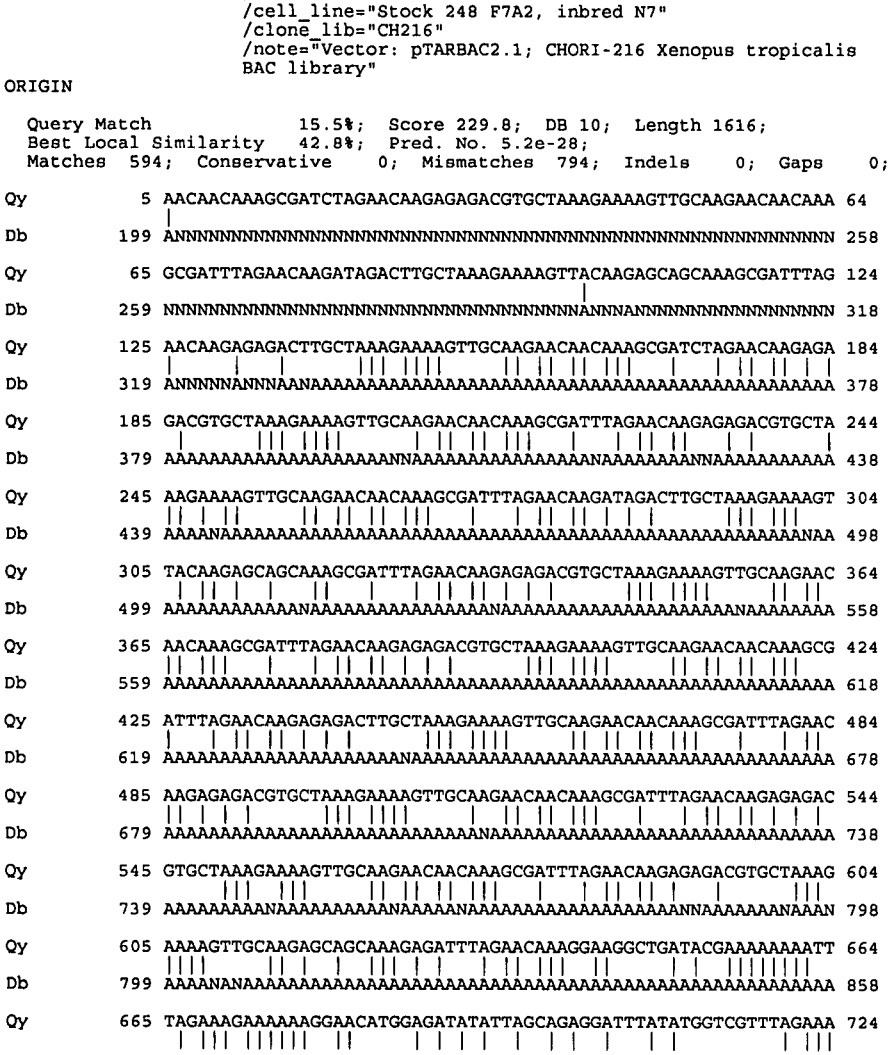


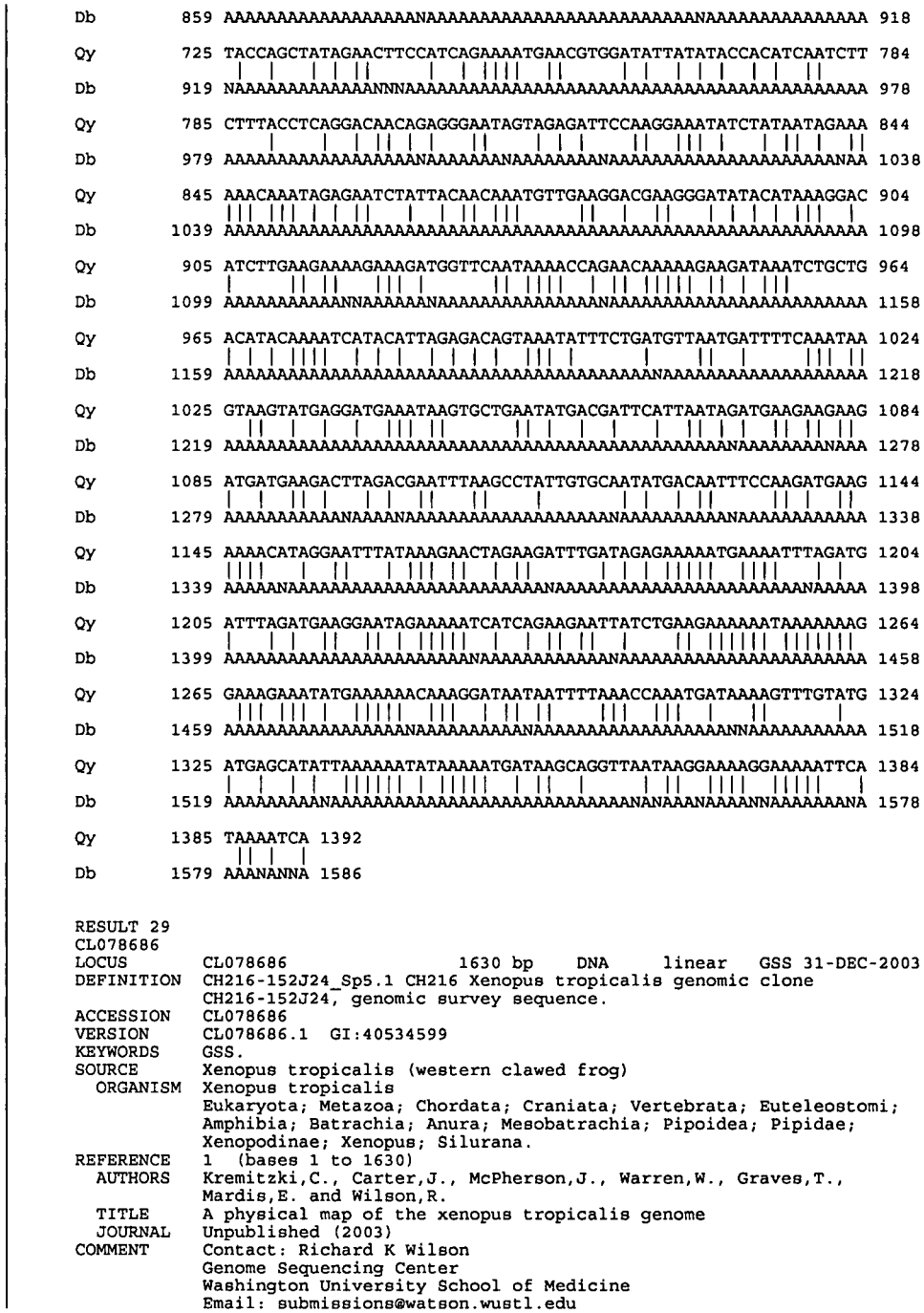


```
RESULT 28
CL081995
LOCUS       CL081995                1616 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-165M19_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-165M19, genomic survey sequence.
ACCESSION   CL081995
VERSION     CL081995.1  GI:40537908
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
```

```
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 1038
            High quality sequence stop: 1108.
FEATURES             Location/Qualifiers
     source          1. .1616
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-165M19"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN
```

Query Match 15.5%; Score 229.8; DB 10; Length 1616;
Best Local Similarity 42.8%; Pred. No. 5.2e-28;
Matches 594; Conservative 0; Mismatches 794; Indels 0; Gaps 0;





```
RESULT 29
CL078686
LOCUS       CL078686               1630 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-152J24_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-152J24, genomic survey sequence.
ACCESSION   CL078686
VERSION     CL078686.1  GI:40534599
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1630)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
```

```
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 498
            High quality sequence stop: 596.
FEATURES             Location/Qualifiers
     source          1. .1630
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-152J24"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match           15.5%;  Score 229.6;  DB 10;  Length 1630;
  Best Local Similarity   48.8%;  Pred. No. 5.6e-28;
  Matches  678;  Conservative    0;  Mismatches  704;  Indels    6;  Gaps    2;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
              || || |  |||    |   | || || | | | |    ||  ||||   | || || ||
Db        216 AAAAAAAGGAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAGAAAAAAAAGAAAAAAAAA 275

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
              |||    |   | | ||  | |||       |||  ||||  | ||| |  || |||   ||
Db        276 AAAAAAAGAGAAATCATAAAAGAAAAAGAAAAAGAAAGAAAAAAAAAAAGAAAA---ATA 332

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
              | || || |  ||     ||| ||||    || || || | |    |   | || ||
Db        333 AAAAAAAAAAAGGAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAA 392

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
              | | |       ||| ||||     || || || |||    |   | || || | | |
Db        393 AAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAA 452

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
              | ||| |||||     || || || |||     |    | || || | | |       ||| |||
Db        453 CAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAA 512

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
              |   | || |  | ||||  |   | || || | | |      ||| ||||     ||
Db        513 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAA 572

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
              || || ||||  |   | || || | | ||     ||| ||||     || || || |||
Db        573 AAAAAAAAAGAAAAAAAAAAAAAAAAACACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 632

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
                |   || || | | |   | ||| ||||    | || || || |||   |   ||
Db        633 AAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAAAAAAAAAA---AAAAAG 689

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
              || || | | |      ||||||||     || || || |||       | || || | |
Db        690 AAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAGAAAAAACAAAAAAAAAAAAAA 749

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
              |      ||| |||||    || || || |||    |   | || || | | |      |
Db        750 AAAAAAAAAAAAAAAAGAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAGA 809

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
              || ||||     || |  |   || | |   |||| |||  ||    || |  |||||||
Db        810 AAAAAAAAAAAAAAAAAAAAACAAAAAAAAGAGAAAAAAAAAAAAAAAGAAAAAAAAAAA 869

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
              |   | ||||||||||   |  |   | | |  || | |   |  | | |    |   |
Db        870 AAAAAAAAAGAAAAAAAACAGAAAAAAACAGAAAAGAAAAAAAAAAGAAAAGAAAGAAAAA 929

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
              |   |  |   | | ||     |  | ||||  ||     | |  | | |  | |  ||

Db        930 ACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 989

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAG 841
                    |   | || || | |   || | ||| |    ||  ||| |   | || |
Db        990 AAAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAGAAAAA 1049

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
              ||||| |||   | ||   |  | || |||     || ||  ||   | | | | |||
Db       1050 AAAAAGAAAAAGAAAAAAAAAGAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAGAAAAAA 1109

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
              | |     || |||| ||| |      |  |||  ||||  ||||||||| | |
Db       1110 AAAAAAAAAAAAAAAAAAAAAACAAAAACAAGAAAGAAGAAAGAAAAGAAGAAAGAAAAA 1169

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
                | |   | ||  | | |  | | | |  ||| |      |    ||  |     |||
Db       1170 AAAAAAAGAAGAAAGAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAA 1229

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
                |  ||     ||  |  ||  ||       ||| |  |  |  |  |  |    | ||
Db       1230 AAGAAAAAAGAGAAAAAAAAGGAAAGAAAAAAAACAAAAAAAAAAAGAAAAGAAAGAAAA 1289

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
              || |  |  ||||   |||  ||   ||     |      || |  | ||    || ||
Db       1290 AAAAAAAGAAAGAACAAGAAAAAGAAAAAGAAAAAAAAGAAAAGAAGAAACAACAGGAGA 1349

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
              || |||| | || ||    | | | |||   ||||     | | | ||| |  ||||   |
Db       1350 AAAAAAAAACAGAAAAAGAAAGAAAACACAAAGAAGAAAAAAAAAAAAGAGAAAAAAAAAA 1409

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
              | ||   |||  ||  || | |||          || ||   |   |||||||| || ||
Db       1410 AAGAAAAAGAGAAATAAAAAAAAAGGAAGAGGAAAAAAAAAAAGAAAGAAAAAAAAAGAA 1469

Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
              ||| ||  ||| |  | |||| ||   | |  ||      ||  ||  ||  |||
Db       1470 AAGAAACAAAAAAAAACAAAAGAACAAAAACAAACGAAGAAAGAAGAGAAGAAAAAAGAA 1529

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAAT 1381
              | | ||  | |  ||||||   |||||  | ||| |    || ||| ||||   | |||
Db       1530 AAAAAGAAAAAAAGAAAAAAAGAAAAAAAAAAAAAGAAAAGAAAGAAGAAAAAAAAGAAAA 1589

Qy       1382 TCATAAAA 1389
                | ||||
Db       1590 GAAAAAAA 1597


RESULT 30
CL082658
LOCUS       CL082658               1457 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-169B13_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-169B13, genomic survey sequence.
ACCESSION   CL082658
VERSION     CL082658.1  GI:40538571
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1457)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 393
```

```
          High quality sequence stop: 1347.
FEATURES             Location/Qualifiers
     source          1. .1457
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-169B13"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.5%;  Score 229.2;  DB 10;  Length 1457;
  Best Local Similarity   48.2%;  Pred. No. 6.6e-28;
  Matches  626;  Conservative    0;  Mismatches  669;  Indels    4;  Gaps    1;

Qy         91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 150
              ||| |||                                | |  | | |       ||| ||||
Db        149 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAANAAAAAAAAANAAAAAAAAAA 208

Qy        151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
                 || || || |||    | | | || || | | |       ||| | ||    | |||||
Db        209 AAAAAAAAAAAAAAAAAAAACCAAAAAAAAAAAAAAAAANAAAAAAAAAGAAAAAGAAAGAA 268

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
               || |||  ||   | || || | | |     |  || ||||      || || || |||
Db        269 AAAAAAAAAGAAAAAAAAAAAAAAAAAAAAACAGAAAAAAAAAAAAAAAAAAAAAAAAAA 328

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
               |   | || || | | |       ||| ||||   | || |  |  |||   |   | ||
Db        329 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 388

Qy        331 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
               || | | |      ||| ||||     || || || |||   |   | || || | | |
Db        389 AAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 448

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAA 450
                 ||| ||||     || || || |||   |     || || | | |      |||
Db        449 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAA 508

Qy        451 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
               |||||    || || || |||   |   | || || | | |      || |||||
Db        509 AAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGGAAAAAAAA 568

Qy        511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
               || || || |||  ||   | || || | | |      ||| ||||     || || ||
Db        569 AAAAAAAAAAAAAAGGAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 628

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
               |||       | || || | | |       ||| ||||   | || |  |  | | | |
Db        629 AAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAGAAAAAA 688

Qy        631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
                 | || |||  ||     | | |||||||||   | ||| ||||||  || |   | |
Db        689 AAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 748

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
               | |  |  | |  |   | |         | ||| |  |   | | ||     |  | ||
Db        749 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA----AAAAAAAA 804

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
               ||  ||   | | |  | | |  | |  ||        |    |  | || | |    ||
Db        805 AAAAAAAACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 864

Qy        811 AGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
               |  | | |    ||  ||| |   | || | |||||| ||| | | ||   |  | || |
Db        865 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 924

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930
               || |   ||  |  ||   | | | | |||  | |     || |||| ||| |      |

Db        925 AAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 984

Qy        931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
               | ||||  | || |||||||| | |||       | | | ||||  | | |  | | | |
Db        985 AAAAAAAAACAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1044

Qy        991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCT 1050
               | ||| |           ||  |     ||| ||| ||  |  |  |  ||| ||
Db       1045 GAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAA 1104

Qy       1051 GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAG 1110
                 | |  |  |    |  || | |  || || || |  |  || |   | |  ||   ||
Db       1105 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1164

Qy       1111 CCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTA 1170
                  |      || |  | ||     || |  || |||| | |  ||   | ||| ||  |
Db       1165 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1224

Qy       1171 GAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCA 1230
                 | ||    | | |  |||| |||   | |  |   | |  ||  || | |||||  |
Db       1225 AACGAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1284

Qy       1231 TCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT 1290
                 | || ||  |    || |||||| |||||||| ||| ||| |  ||||||  |||  |
Db       1285 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAA 1344

Qy       1291 AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT 1350
                 || |     |||  |||  | ||||      |  |  |  | |  ||||||   |||||
Db       1345 AAAAGAAAAAAAAGAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAA 1404

Qy       1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
                 | ||  |    || ||  ||||  |||||    | ||||
Db       1405 AAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1443


Search completed: December 20, 2005, 15:07:52
Job time : 7889.23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:28:07 ; Search time 372.484 Seconds
(without alignments)
5614.164 Million cell updates/sec

Title: US-09-837-344-41
Perfect score: 2486
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22916/app_query.fasta_1.3740
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_1319_@runat_21122005_093215_22916 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2478	99.7	493	2	Q25886_PLAFA	Q25886 plasmodium
2	2406.5	96.8	1909	2	Q25893_PLAFA	Q25893 plasmodium
3	2375.5	95.6	1596	2	Q8IJ44_PLAF7	Q8ij44 plasmodium
4	1451	58.4	280	2	Q25844_PLAFA	Q25844 plasmodium
5	1450	58.3	280	2	Q25843_PLAFA	Q25843 plasmodium
6	1447	58.2	280	2	Q25847_PLAFA	Q25847 plasmodium
7	1447	58.2	280	2	Q25848_PLAFA	Q25848 plasmodium
8	1447	58.2	280	2	Q25852_PLAFA	Q25852 plasmodium
9	1445	58.1	280	2	Q25849_PLAFA	Q25849 plasmodium
10	1445	58.1	280	2	Q25851_PLAFA	Q25851 plasmodium
11	1443	58.0	280	2	Q25850_PLAFA	Q25850 plasmodium
12	1442	58.0	280	2	Q25846_PLAFA	Q25846 plasmodium
13	1437	57.8	280	2	Q25853_PLAFA	Q25853 plasmodium
14	1434	57.7	280	2	Q25854_PLAFA	Q25854 plasmodium
15	1433	57.6	280	2	Q25845_PLAFA	Q25845 plasmodium
16	1432	57.6	280	2	Q25900_PLAFA	Q25900 plasmodium
17	1430	57.5	280	2	Q25855_PLAFA	Q25855 plasmodium
18	1430	57.5	280	2	Q25888_PLAFA	Q25888 plasmodium
19	1428	57.4	280	2	Q27243_PLAFA	Q27243 plasmodium
20	1426	57.4	280	2	Q25889_PLAFA	Q25889 plasmodium
21	1366	54.9	264	2	Q9GTX5_PLAFA	Q9gtx5 plasmodium
22	770.5	31.0	318	2	Q25887_PLAFA	Q25887 plasmodium
23	468.5	18.8	978	2	Q7RAK4_PLAYO	Q7rak4 plasmodium
24	451	18.1	1207	2	Q4UHB4_THEAN	Q4uhb4 theileria a
25	432	17.4	905	2	Q54XN9_DICDI	Q54xn9 dictyosteli
26	429.5	17.3	3427	2	Q6YA79_PLARE	Q6ya79 plasmodium
27	428	17.2	1082	2	Q54JG8_DICDI	Q54jg8 dictyosteli
28	422	17.0	2205	2	Q54C75_DICDI	Q54c75 dictyosteli
29	419	16.9	762	2	Q50RA3_ENTHI	Q50ra3 entamoeba h
30	419	16.9	1061	2	Q54MJ1_DICDI	Q54mj1 dictyosteli
31	418	16.8	1575	2	Q519G4_ENTHI	Q519g4 entamoeba h
32	411	16.5	1738	2	Q51ED7_ENTHI	Q51ed7 entamoeba h
33	403	16.2	1119	2	Q54TT8_DICDI	Q54tt8 dictyosteli
34	398.5	16.0	1108	2	Q9NDI0_BABBI	Q9ndi0 babesia big
35	393.5	15.8	1508	2	Q75JP5_DICDI	Q75jp5 dictyosteli
36	391.5	15.7	2349	2	Q8I455_PLAF7	Q8i455 plasmodium
37	391	15.7	1750	2	Q8X0H2_NEUCR	Q8x0h2 neurospora
38	389	15.6	826	2	Q9VPS3_DROME	Q9vps3 drosophila
39	383.5	15.4	1777	2	Q54FB8_DICDI	Q54fb8 dictyosteli
40	381.5	15.3	1003	2	Q91LX9_HHV8	Q91lx9 human herpe
41	381.5	15.3	1129	2	Q9QR71_HHV8	Q9qr71 human herpe
42	380.5	15.3	1407	1	TRHY_RABIT	P37709 oryctolagus
43	380.5	15.3	1549	1	TRHY_SHEEP	P22793 ovis aries
44	377.5	15.2	1089	2	O40947_HHV8	O40947 human herpe
45	377	15.2	1036	2	Q9DUM3_HHV8	Q9dum3 human herpe

ALIGNMENTS

RESULT 1
Q25886_PLAFA
ID Q25886_PLAFA PRELIMINARY; PRT; 493 AA.
AC Q25886;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage antigen-1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94267224; PubMed=7515922;
RA Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT "Plasmodium falciparum liver stage antigen-1 is well conserved and
RT contains potent B and T cell determinants.";
RL J. Immunol. 153:190-204(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT "The Plasmodium falciparum liver stage antigen LSA-1 is well conserved
RT and harbors major B- and T-cell epitopes.";
RL J. Immunol. 0:0-0(1994).
DR EMBL; Z30320; CAA82975.1; -; Genomic_DNA.
FT NON_TER 1 1
SQ SEQUENCE 493 AA; 59062 MW; FE252CC44F46996E CRC64;

Alignment Scores:

Pred. No.:	6.07e-108	Length:	493
Score:	2478.00	Matches:	493
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.68%	Indels:	0
DB:	2	Gaps:	0

```
US-09-837-344-41 (1-1482) x Q25886_PLAFA (1-493)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 20

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 40

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 60

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 80

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGlu 100

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGln 120

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 140

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLys 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu 240

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 260

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 280

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 300

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 320

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGln 340

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        341 IleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGlu 360

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAsp 380

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeu 400

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLys 420

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeu 440

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 TyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLys 460

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 PheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAsp 480

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||
Db        481 GluLeuSerGluAspIleThrLysTyrPheMetLysLeu 493

RESULT 2
Q25893_PLAFA
ID   Q25893_PLAFA PRELIMINARY;      PRT;  1909 AA.
AC   Q25893;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Liver stage antigen.
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NF54;
RX   MEDLINE=92107224; PubMed=1840628; DOI=10.1016/0166-6851(91)90117-O;
RA   Zhu J., Hollingdale M.R.;
RT   "Structure of Plasmodium falciparum liver stage antigen-1.";
RL   Mol. Biochem. Parasitol. 48:223-226(1991).
DR   EMBL; X56203; CAA39663.1; -; Genomic_DNA.
DR   PIR; S24597; A45592.
DR   HSSP; O15813; 1D7M.
SQ   SEQUENCE   1909 AA;  230153 MW;  8B861A524C797217 CRC64;

Alignment Scores:
Pred. No.:              9.77e-105       Length:         1909
Score:                  2406.50         Matches:        487
Percent Similarity:     89.89%          Conservative:   2
Best Local Similarity:  89.52%          Mismatches:     4
Query Match:            96.80%          Indels:         51
DB:                     2               Gaps:           2

US-09-837-344-41 (1-1482) x Q25893_PLAFA (1-1909)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT------------------------------- 87
              |||||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405

Qy         88 ------------------GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
```

```
Db      1406 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 1425

Qy       130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189

Db      1426 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 1445

Qy       190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249

Db      1446 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 1465

Qy       250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 309

Db      1466 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 1485

Qy       310 GAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 369

Db      1486 GluGlnGlnArgAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 1505

Qy       370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429

Db      1506 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 1525

Qy       430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489

Db      1526 GluGlnGluArgLeuAlaAsnGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnGlu 1545

Qy       490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549

Db      1546 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1565

Qy       550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------------ 597

Db      1566 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1585

Qy       597 ------------------------------------------------------------ 597

Db      1586 LeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlu 1605

Qy       598 ----------------------------GCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627

Db      1606 GlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArg 1625

Qy       628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687

Db      1626 AspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGly 1645

Qy       688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747

Db      1646 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 1665

Qy       748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807

Db      1666 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 1685

Qy       808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867

Db      1686 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 1705

Qy       868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927

Db      1706 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 1725

Qy       928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987

Db      1726 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 1745

Qy       988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047

Db      1746 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 1765

Qy      1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107

Db      1766 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 1785

Qy      1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167

Db      1786 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 1805

Qy      1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227

Db      1806 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 1825

Qy      1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287

Db      1826 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 1845

Qy      1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347

Db      1846 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 1865

Qy      1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407

Db      1866 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 1885

Qy      1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467

Db      1886 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 1905

Qy      1468 TTTATGAAACTA 1479

Db      1906 PheMetLysLeu 1909
```

RESULT 3
Q8IJ44_PLAF7
ID Q8IJ44_PLAF7 PRELIMINARY; PRT; 1596 AA.
AC Q8IJ44;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Liver stage antigen, putative.
GN ORFNames=PF10_0356;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35553.1; -; Genomic_DNA.
SQ SEQUENCE 1596 AA; 191691 MW; C0B3A2CC0F5D60AC CRC64;

Alignment Scores:

Pred. No.:	2.82e-103	Length:	1596
Score:	2375.50	Matches:	478
Percent Similarity:	94.71%	Conservative:	5
Best Local Similarity:	93.73%	Mismatches:	10
Query Match:	95.56%	Indels:	17
DB:	2	Gaps:	1

US-09-837-344-41 (1-1482) x Q8IJ44_PLAF7 (1-1596)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db       1087 GlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGln 1106
```

```
Qy        61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
             ||||||||||||||||||:::|||||||||||||||||||||||||   ||||||||||||
Db      1107 GlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnGlnSerAsp 1126

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
             ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db      1127 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnGlnSerAspLeuGluGln 1146

Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
             ||||||   ||||||||||||||||||||:::||||||||||||||||||||||||||||
Db      1147 GluArgLeuAlaLysGluLysLeuGlnGluArgGlnSerAspLeuGluGlnGluArgLeu 1166

Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
             ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db      1167 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGlu 1186

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
             ||||||||||||||||||||||||||||||||:::|||   ||||||||||||||||||||
Db      1187 LysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGln 1206

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG---------------- 405
             |||||||||||||||||||||||||||||||   ||||||||||||
Db      1207 GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnHis 1226

Qy       406 ----------------------------------TTGCAAGAACAACAAAGCGATTTA 429
                                                ||||||||||||||||||||||||
Db      1227 SerAspLeuGluArgThrLysAlaSerLysGluThrLeuGlnGluGlnGlnSerAspLeu 1246

Qy       430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1247 GluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 1266

Qy       490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1267 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1286

Qy       550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1287 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1306

Qy       610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1307 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 1326

Qy       670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db      1327 ArgLysLysGluHisGlyAspValLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 1346

Qy       730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1347 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 1366

Qy       790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACA 849
             ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      1367 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluAsnThr 1386

Qy       850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1387 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 1406

Qy       910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1407 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSerAlaAspIle 1426

Qy       970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1427 GlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGlnIleSerLys 1446

Qy      1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1447 TyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGluGluAspAsp 1466

Qy      1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1467 GluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAspGluGluAsn 1486

Qy      1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1487 IleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeu 1506

Qy      1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1507 AspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLysLysGlyLys 1526

Qy      1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1527 LysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGlu 1546

Qy      1330 CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1547 HisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLysPheIleLys 1566

Qy      1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1567 SerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSer 1586

Qy      1450 GAAGATATAACTAAATATTTTATGAAACTA 1479
             |||||||||||||||||||||||||||||
Db      1587 GluAspIleThrLysTyrPheMetLysLeu 1596
```

RESULT 4
Q25844_PLAFA
ID Q25844_PLAFA PRELIMINARY; PRT; 280 AA.
AC Q25844;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; L40885; AAB59231.1; -; Genomic_DNA.
FT NON_TER 1 1
SQ SEQUENCE 280 AA; 32940 MW; E9708E3CFFAEA9CF CRC64;

Alignment Scores:

Pred. No.:	4.62e-60	Length:	280
Score:	1451.00	Matches:	280
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	58.37%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-41 (1-1482) x Q25844_PLAFA (1-280)

```
Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy      1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 5
Q25843_PLAFA
ID   Q25843_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25843;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40884; AAB59230.1; -; Genomic_DNA.
FT   NON_TER      1      1
SQ   SEQUENCE   280 AA;  32926 MW;  467080F32FAEAD33 CRC64;

Alignment Scores:
Pred. No.:              5.15e-60       Length:        280

Score:                  1450.00        Matches:       279
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  99.64%         Mismatches:    0
Query Match:            58.33%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-41 (1-1482) x Q25843_PLAFA (1-280)

Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy      1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 6
Q25847_PLAFA
ID   Q25847_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25847;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
```

```
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40888; AAC41598.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32944 MW;  467085E23BABBC27 CRC64;

Alignment Scores:
Pred. No.:              7.1e-60        Length:         280
Score:                  1447.00        Matches:        278
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  99.29%         Mismatches:     0
Query Match:            58.21%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25847_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysMetLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 7
Q25848_PLAFA
ID   Q25848_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40889; AAC41599.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32939 MW;  E77080385FAA0D61 CRC64;

Alignment Scores:
Pred. No.:              7.1e-60        Length:         280
Score:                  1447.00        Matches:        279
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  99.64%         Mismatches:     0
Query Match:            58.21%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25848_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 8
Q25852_PLAFA
ID   Q25852_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25852;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40834; AAB59233.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32927 MW;  E6648F85FBBAA245 CRC64;

Alignment Scores:
Pred. No.:              7.1e-60        Length:        280
Score:                  1447.00        Matches:       278
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  99.29%         Mismatches:    0
Query Match:            58.21%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-41 (1-1482) x Q25852_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                 ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                 |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 9
Q25849_PLAFA
ID   Q25849_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40890; AAC41600.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32966 MW;  E96D255154DEA9CF CRC64;

Alignment Scores:
Pred. No.:              8.8e-60        Length:        280
Score:                  1445.00        Matches:       279
```

```
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  99.64%         Mismatches:    0
Query Match:            58.13%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-41 (1-1482) x Q25849_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluTyrIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 10
Q25851_PLAFA
ID   Q25851_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40893; AAC41603.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32927 MW;  48706EFDCF40ADD3 CRC64;

Alignment Scores:
Pred. No.:              8.8e-60        Length:        280
Score:                  1445.00        Matches:       278
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  99.29%         Mismatches:    0
Query Match:            58.13%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-41 (1-1482) x Q25851_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAspGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240
```

```
Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 11
Q25850_PLAFA
ID   Q25850_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25850;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40891; AAC41601.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32989 MW;  E96F812CFFAEBED8 CRC64;

Alignment Scores:
Pred. No.:              1.09e-59       Length:         280
Score:                  1443.00        Matches:        279
Percent Similarity:     99.64%         Conservative:   0
Best Local Similarity:  99.64%         Mismatches:     1
Query Match:            58.05%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25850_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspTyrAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 12
Q25846_PLAFA
ID   Q25846_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40887; AAC41597.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32882 MW;  46763641C81AFC33 CRC64;

Alignment Scores:
Pred. No.:              1.21e-59       Length:         280
Score:                  1442.00        Matches:        278
Percent Similarity:     99.64%         Conservative:   1
Best Local Similarity:  99.29%         Mismatches:     1
Query Match:            58.00%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25846_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
```

```
Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 13
Q25853_PLAFA
ID   Q25853_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25853;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40835; AAB59234.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32938 MW;  46751C45F4DCBD33 CRC64;

Alignment Scores:
Pred. No.:              2.08e-59        Length:         280
Score:                  1437.00         Matches:        277
Percent Similarity:     99.29%          Conservative:   1
Best Local Similarity:  98.93%          Mismatches:     2
Query Match:            57.80%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-41 (1-1482) x Q25853_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 14
Q25854_PLAFA
ID   Q25854_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25854;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40836; AAC41604.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32966 MW;  61851C5AD6E1A211 CRC64;

Alignment Scores:
Pred. No.:              2.86e-59        Length:         280
Score:                  1434.00         Matches:        276
Percent Similarity:     99.29%          Conservative:   2
Best Local Similarity:  98.57%          Mismatches:     2
Query Match:            57.68%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-41 (1-1482) x Q25854_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluArgIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 15
Q25845_PLAFA
ID   Q25845_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25845;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40886; AAB59232.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32908 MW;  FDEE929F34DCA7E9 CRC64;

Alignment Scores:
Pred. No.:              3.19e-59        Length:         280
Score:                  1433.00         Matches:        276
Percent Similarity:     98.93%          Conservative:   1
Best Local Similarity:  98.57%          Mismatches:     3
Query Match:            57.64%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-41 (1-1482) x Q25845_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160
```

```
Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 16
Q25900_PLAFA
ID   Q25900_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25900;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40947; AAC41592.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32911 MW;  F8850CABF6136212 CRC64;

Alignment Scores:
Pred. No.:              3.55e-59       Length:         280
Score:                  1432.00        Matches:        276
Percent Similarity:     99.29%         Conservative:   2
Best Local Similarity:  98.57%         Mismatches:     2
Query Match:            57.60%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25900_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValSerAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 17
Q25855_PLAFA
ID   Q25855_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25855;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40837; AAC41605.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32866 MW;  73164A709789DD31 CRC64;

Alignment Scores:
Pred. No.:              4.4e-59        Length:         280
Score:                  1430.00        Matches:        276
Percent Similarity:     98.93%         Conservative:   1
Best Local Similarity:  98.57%         Mismatches:     3
```

```
Query Match:          57.52%      Indels:        0
DB:                   2           Gaps:          0

US-09-837-344-41 (1-1482) x Q25855_PLAFA (1-280)

Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
             |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGlyAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy      1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 18
Q25888_PLAFA
ID   Q25888_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25888;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40908; AAB59228.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32936 MW;  FDF9A29104C5A7E9 CRC64;

Alignment Scores:
Pred. No.:            4.4e-59      Length:        280
Score:                1430.00      Matches:       275
Percent Similarity:   98.93%       Conservative:  2
Best Local Similarity: 98.21%      Mismatches:    3
Query Match:          57.52%       Indels:        0
DB:                   2            Gaps:          0

US-09-837-344-41 (1-1482) x Q25888_PLAFA (1-280)

Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
             |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheArgProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 19
Q27243_PLAFA
ID   Q27243_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q27243;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40910; AAC41582.1; -; Genomic_DNA.
DR   EMBL; L40892; AAC41602.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32881 MW;  105849F296958C78 CRC64;

Alignment Scores:
Pred. No.:              5.45e-59       Length:         280
Score:                  1428.00        Matches:        275
Percent Similarity:     98.93%         Conservative:   2
Best Local Similarity:  98.21%         Mismatches:     3
Query Match:            57.44%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q27243_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluSerIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 20
Q25889_PLAFA
ID   Q25889_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40909; AAB59229.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32850 MW;  FDEE929F34DF68E9 CRC64;

Alignment Scores:
Pred. No.:              6.75e-59       Length:         280
Score:                  1426.00        Matches:        275
Percent Similarity:     98.57%         Conservative:   1
Best Local Similarity:  98.21%         Mismatches:     4
Query Match:            57.36%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-41 (1-1482) x Q25889_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValGlyGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 21
Q9GTX5_PLAFA
ID   Q9GTX5_PLAFA PRELIMINARY;      PRT;   264 AA.
AC   Q9GTX5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Xu J., Chen S.Y.;
RT   "Cloning and sequence analysis of 3' region of liver stage antigen-1
RT   gene of Plasmodium falciparum isolate FCC1/HN.";
RL   Zhongshan Yi Ke Da Xue Xue Bao 22:410-413(2001).
DR   EMBL; AF246996; AAG12324.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   264 AA;  31006 MW;  1714D653E8D9D1D7 CRC64;

Alignment Scores:
Pred. No.:              4.26e-56        Length:         264
Score:                  1366.00         Matches:        263
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.62%          Mismatches:     0
Query Match:            54.95%          Indels:         0
DB:                     2               Gaps:           0
```

```
US-09-837-344-41 (1-1482) x Q9GTX5_PLAFA (1-264)

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
                 |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 20

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 40

Qy        808 AATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 60

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 80

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 100

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 120

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 140

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 160

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 180

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 200

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 220

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 240

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 260

Qy       1468 TTTATGAAACTA 1479
                 ||||||||||||
Db        261 PheMetLysLeu 264

RESULT 22
Q25887_PLAFA
ID   Q25887_PLAFA PRELIMINARY;      PRT;   318 AA.
AC   Q25887;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Liver stage antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=94267224; PubMed=7515922;
```

```
RA   Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA   Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA   Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT   "Plasmodium falciparum liver stage antigen-1 is well conserved and
RT   contains potent B and T cell determinants.";
RL   J. Immunol. 153:190-204(1994).
DR   EMBL; Z30319; CAA82974.1; -; Genomic_DNA.
DR   HSSP; O15813; 1D7M.
FT   NON_TER     318     318
SQ   SEQUENCE   318 AA;  38098 MW;  989F32A70605DFB2 CRC64;

Alignment Scores:
Pred. No.:              2.22e-28       Length:         318
Score:                  770.50         Matches:        163
Percent Similarity:     92.59%         Conservative:   12
Best Local Similarity:  86.24%         Mismatches:     13
Query Match:            30.99%         Indels:         1
DB:                     2              Gaps:           1

US-09-837-344-41 (1-1482) x Q25887_PLAFA (1-318)

Qy         16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 75
              |||   ::::::::::::      |||   :::   :::   :::|||   |||||||||
Db        131 AspAspAspLysLysLysTyrIleLysGlyGlnAspGluAsnArgGlnGluAspLeuGlu 150

Qy         76 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 135
                ::::::   ||||||||||||||||||   ||||||||||||   ||||||||||||
Db        151 ---GluLysAlaAlaLysGluLysLeuGlnGlyGlnGlnSerAspSerGluGlnGluArg 169

Qy        136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        170 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLys 189

Qy        196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 209

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        210 GlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 229

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAsp 249

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 269

Qy        436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        270 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeu 289

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 309

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTA 582
              :::|||||||||||||||||||||||
Db        310 ArgLeuGlnGluGlnGlnSerAspLeu 318

RESULT 23
Q7RAK4_PLAYO
ID   Q7RAK4_PLAYO PRELIMINARY;      PRT;   978 AA.
AC   Q7RAK4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Glutamine-asparagine rich protein.
GN   Name=PY06496;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01002201; EAA18721.1; -; Genomic_DNA.
SQ   SEQUENCE   978 AA;  114234 MW;  B72DC75FDDEB7A27 CRC64;

Alignment Scores:
Pred. No.:              2.04e-14       Length:         978
Score:                  468.50         Matches:        152
Percent Similarity:     57.14%         Conservative:   124
Best Local Similarity:  31.47%         Mismatches:     134
Query Match:            18.85%         Indels:         73
DB:                     2              Gaps:           23

US-09-837-344-41 (1-1482) x Q7RAK4_PLAYO (1-978)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||   :::|||      ||||||::::::|||      :::|||||||||:::::::::
Db        318 GlnSerGluGlnGluLysLeuGluGluLysArgLeuGluGluGluLysLeuLysGlnGlu 337

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::      |||:::|||::::::||||||:::|||||||||   :::::::::
Db        338 Lys------LeuLysGlnGluLysLeuAlaGlnGluLysLeuAlaGlnGluArg------ 353

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||   ||||||:::||||||:::|||||||||:::::::::|||      ||||||:::
Db        354 LeuAlaGlnGluLysLeuAlaGlnGluLysLeuLysGlnGluGln------LeuGluArg 371

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||:::   |||:::|||||||||:::::::::::      |||:::||||||:::
Db        372 GluLysLeuAlaGlnGluLysLeuLysGlnGluLys------LeuLysGlnGluGlnLeu 389

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||:::||||||                  |||:::|||:::::|||   :::|||
Db        390 AlaGlnGluLys--------------------LeuLysGlnGluLysLeuLysGlnGlu 402

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||::::::::::|||      ||||||:::||||||   |||:::||||||||:::
Db        403 LysLeuLysGlnGluGln------LeuGluArgGluArgLeuAlaGlnGluLysLeuLys 420

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ::::::|||    ||||||||||||||   |||:::||||||||:::::::::|||
Db        421 GlnGluGln------LeuGluGlnGluArgLeuAlaGlnGluLysLeuLysGlnGluGln 438

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||:::||||||||||:::||||||||:::::::::|||      |||
Db        439 ------LeuGluArgGluArgLeuAlaGlnGluLysLeuLysGlnGluGln------Leu 454

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------CAAAGCGATTTA 531
              ||||||||||||   |||:::|||||||||:::::::::         |||      |||
Db        455 GluGlnGluArgLeuAlaGlnGluLysLeuLysGlnGluArgLeuAlaGlnGluLysLeu 474

Qy        532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTA 582
              |||:::|||:::      :::|||||||||:::   ::::::   :::         :::
Db        475 GluArgGluLysLeuGluArgGluLysLeuGluArgGluLysIleGluArgGluLysIle 494
```

```
Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
Db        495 GluArgGluLysIleGluArgGluLysIleGluArgGluGln------LeuGluLysLys 512

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
Db        513 LysLeuGlu---GluLysArgLeuGluGluLysArgLeuGluGluLysArgLeuGluGlu 531

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
Db        532 Lys---------ArgLeuGluGluLysArgIleGluGluGluLysArgIleLysGluGlu 548

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA---GGGAATAGTAGAGAT 819
Db        549 GlnArgIleLysGluGluGlnArgIleGluGluGluLysArgIleLysGluGluLysArg 568

Qy        820 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGA------GAATCTATTACAACAAAT 873
Db        569 IleLysGluGluGlnArgIleGluGluGluLysArgIleLysGluGluLysArgIleLys 588

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db        589 GluGluLysArgIleLysGluGluLysArgIleLysGluGluLysArg---------Ile 605

Qy        934 AAACCAGAACAAAAA------GAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
Db        606 GluGluGluLysLysLeuGluLysGluLysLysProGluLeuGluAsnLysGluArgGly 625

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
Db        626 ThrIleHis-----------------GlnValLeuSerLysIleLysGluLysLeuGln 639

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAA------GAAGATGATGAAGACTTAGAC 1101
Db        640 ProGluSerGlyLysValValGluGlyGluLysLysValGluAlaGlyLysValValAsp 659

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        660 AlaGlyLysValVal----------------------------AspAlaGlyLysVal 669

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT----------------- 1203
Db        670 ValGluAlaGlyLysValValAspAlaGlyLysValValAspAlaGlyLysValValGlu 689

Qy       1204 --------------GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
Db        690 AlaGlyLysValValGluAlaGlyLysValValGluAlaGlyLysValValGluGlyGlu 709

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
Db        710 LysLysValGluGluGlyLysLysValGluProGluLysLysGluAsn---GluThrLeu 728

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
Db        729 AspLysAlaLeuSerLysThrGlnGlnGlnGlnLeuGluGlyGluGlnLysValGluGly 748

Qy       1369 GAAAAGGAA 1377
Db        749 GluGlnLys 751
```

RESULT 24
Q4UHB4_THEAN
ID   Q4UHB4_THEAN PRELIMINARY;   PRT; 1207 AA.
AC   Q4UHB4;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=TA20215;
OS   Theileria annulata.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC   Theileria.
OX   NCBI_TaxID=5874;

RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Ankara isolate clone C9;
RA   Pain A., Renauld H., Murphy L., Harris D.A., Quail M.A., Berriman M.,
RA   Hall N., Barrell B.G.;
RT   "The chromosome 1 sequence of Theileria annulata.";
RL   Submitted (APR-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR940347; CAI73525.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1207 AA; 144117 MW; B7D707AF2E1DB7A2 CRC64;

Alignment Scores:

Pred. No.:	1.27e-13	Length:	1207
Score:	451.00	Matches:	151
Percent Similarity:	51.93%	Conservative:	145
Best Local Similarity:	26.49%	Mismatches:	164
Query Match:	18.14%	Indels:	110
DB:	2	Gaps:	25

US-09-837-344-41 (1-1482) x Q4UHB4_THEAN (1-1207)

```
Qy          1 CAAGAAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        272 GlnArgArgGlnArgArgGluGluArgGluArgLysLysGlnGluArgLeuArgLysGln 291

Qy         61 CAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 117
Db        292 GluGluGluArgLeuGluGlnGluArgLeuGluLysGluArgLeuAlaGluGlnGluArg 311

Qy        118 ---GATTTAGAACAA---------------------------------------------- 129
Db        312 LeuAspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGlu 331

Qy        130 ------GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
Db        332 GluThrGluArgLeuLysLysGluArgLeuGluGlnGluArg------LeuGluLysGlu 349

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------------------TTA 225
Db        350 ArgLeuGluLysGluArgLeuArgLysLeuGluGluGluArgLeuGluLysGluArgLeu 369

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT 285
Db        370 GluGlnGluArgLeuGluGlnAlaGlyIleGluGluGluGlnLysArgLeuGluGlnGlu 389

Qy        286 AGACTTGCT-----------------------AAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db        390 ArgLeuAlaIleGluGluGlnGluLysLeuGluLysGluArgIleArgLysGluGluGlu 409

Qy        322 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---GATTTA 378
Db        410 ArgLeuGluGlnGluArgLeuGluLysGluArgLeuAlaGluGlnGluArgLeuAspIle 429

Qy        379 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
Db        430 GluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGluGluThrGlu 449

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 498
Db        450 ArgLeuLysLysGluArgLeuGluGlnGluArg------LeuGluLysGluArgLeuGlu 467

Qy        499 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558
Db        468 LysGluArgLeuGluGlnGlnArgGlnGluGlnGluArgLeuArgLysLeuGluGluArg 487

Qy        559 TTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGTGCTAAAGAAAAGTTG--- 612
Db        488 LeuGluGlnGluArgLeuAlaIleGluGluGlnGluArgLeuGluLysGluArgIleGlu 507

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
Db        508 GlnGluArgIleArgLysLeuGluGluGlnArgLeuGluLysGluArgLeuAlaGlu--- 526
```

```
Qy        673 AAAAAGGAACATGGAGATATA--------------TTAGCAGAGGATTTATATGGTCGT 717
                ||||||      ||||||                |||:::::::::      |||
Db        527 ---LysGluArgLeuAspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArg 545

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                |||                          |||   ||||||
Db        546 Leu-----------------AlaArgGluGluThrGluArg-----------LeuLys 555

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
                :::     |||   |||:::                |||      ||||||
Db        556 LysGluArgLeuGluGlnGluArgLeuGluLysGluArgLeuGluLysGluArgLeuGlu 575

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                  ::::::      :::|||   :::         :::|||   |||      :::
Db        576 GlnGlnArgGlnGluGlnGluArgLeu---ArgLysLeuGluGluArg------LeuGlu 592

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                |||      :::   |||:::::::::          |||         ::::::::::::
Db        593 LysGluArgIleHisGluGluGlnGluArgLeuGluLysGluArgIleGluGlnGluArg 612

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                  ::::::          ||||||          :::::::::
Db        613 IleArgLysLeuGluGluGlnArgLeuGluLysGluArgLeuAlaGluLysGluArgLeu 632

Qy       1018 CAAATA--------------------AGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
                |||                          :::   :::      ::::::   |||
Db        633 AspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGluGlu 652

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
                  :::   |||       |||       :::   |||   |||:::      |||
Db        653 ThrGluArgLeuLysLysGluArgLeuGluGlnGluArgLeuGlu------LysGluArg 670

Qy       1117 GTGCAATATGACAATTTC------------CAAGATGAAGAAAACATAGGAATTTATAAA 1164
                ::::::   :::                |||:::::|||   :::          :::
Db        671 LeuGluLysGluArgLeuGluGlnGlnArgGlnGluGlnGluArgLeu---------Arg 687

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
                :::||||||:::   :::|||||||   |||   :::   :::       |||   :::|||
Db        688 LysLeuGluGluArgLeuGluLys---GluArgIleHisGluGluGlnGluArgLeuGlu 706

Qy       1225 AAATCATCAGAAGAA-----------------TTATCTGAAGAAAAAATAAAAAAAGGA 1266
                |||         |||                   |||   |||:::::::::::|||
Db        707 LysGluArgIleGluGlnGluArgIleArgLysLeuGluGluGlnArgLeuGluLysGlu 726

Qy       1267 AAGAAATATGAAAAA---------------------------------------------- 1281
                :::     ||||||
Db        727 ArgLeuAlaGluLysGluArgLeuAspIleGluGluLysIleArgPheAlaGlnGluVal 746

Qy       1282 ------------ACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
                          ::::::          :::         :::   |||      |||
Db        747 GlnLysArgLeuAlaArgGluGluThrGluArgLeuLysLysGluArgLeuGluGlnGlu 766

Qy       1330 CATATTAAAAAATATAAAAATGATAAGCAG---GTTAATAAGGAAAAGGAAAAATTCATA 1386
                  ::::::|||   :::   :::|||:::   :::   |||::::::|||:::   :::
Db        767 ArgLeuGluLysGluArgLeuGluLysGluArgLeuArgLysGlnGluGluGluArgLeu 786

Qy       1387 AAA--------------------TCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
                :::                  ::::::   :::         ||||||   :::|||
Db        787 GluLysGluArgLeuSerGlyIleAsnValGlyAsnLysArgSerGlyAspTyrAspIle 806

Qy       1426 TTACAGATCGTGGATGAGTTATCTGAAGAT 1455
                  |||      :::|||   |||
Db        807 ProAlaPheValProSerIleSerAlaAsp 816

RESULT 25
Q54XN9_DICDI
ID   Q54XN9_DICDI PRELIMINARY;      PRT;   905 AA.
AC   Q54XN9;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0206221;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000053; EAL68014.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   905 AA;  105156 MW;  AEE07236412B83CC CRC64;

Alignment Scores:
Pred. No.:              1.04e-12        Length:         905
Score:                  432.00          Matches:        146
Percent Similarity:     48.23%          Conservative:   99
Best Local Similarity:  28.74%          Mismatches:     145
Query Match:            17.38%          Indels:         118
DB:                     2               Gaps:           20

US-09-837-344-41 (1-1482) x Q54XN9_DICDI (1-905)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
                :::|||   :::       ||||||   ||||||       ||||||:::|||:::   :::
Db        188 LysGluAsnLysGluArgLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGlu 207

Qy         61 CAAAGCGAT---------TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
                :::   :::         ||||||   :::||||||   ||||||||||||:::   :::
Db        208 ArgLeuGluLysGluLysLeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGlu 227

Qy        112 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                :::       ||||||:::|||||||||       |||:::   ::::::::::::
Db        228 Arg------LeuGluLysGluArgLeuGluAlaGluArgPheGluLysGluArg------ 243

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------- 222
                ||||||   ||||||       ||||||:::|||:::   ::::::   :::
Db        244 LeuGluSerGluArgLeuGluLysGluArgLeuGluAlaGluArgLeuGluLysGluLys 263

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
                ||||||   ||||||       ||||||||||||:::   ::::::       ||||||:::
Db        264 LeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGluArg------LeuGluLys 281

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
                :::||||||       |||:::|||::::::::::::       ||||||   ||||||
Db        282 GluArgLeuGluAlaGluArgLeuGluLysGluArg------LeuGluAlaGluArgLeu 299

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGT 393
                ||||||||||||:::   ::::::   :::         ||||||   ||||||
Db        300 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGluArgLeu 319

Qy        394 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACTT 444
                ||||||||||||:::   ::::::   :::         ||||||   |||||||||
```

```
Db        320 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGluArgLeu 339

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGT 495
              |||||||||||||:::   ::::::   :::         ||||||   ||||||
Db        340 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGluArgLeu 359

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
              |||||||||||||:::   ::::::      ||||||:::||||||          |||
Db        360 GluLysGluLysLeuGluAlaGluArg------LeuGluLysGluArgIleGluAsnGlu 377

Qy        556 AAGTTGCAA---GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              |||:::   |||:::         ||||||  |||:::         |||:::|||
Db        378 IleLeuGluAlaGluArgLeuGluArgLeuGluMetGluLysIleGluIleGluArgLeu 397

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ::::::|||:::         ||||||  |||           |||  ||||||:::
Db        398 GluLysGlnArgLeuGlnGluGluGlnIleLysGluLysGluMetLysArgLeuGluLys 417

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              :::|||                                 |||||||||
Db        418 GluLys-----------------------------------ArgLeuGlu--------- 422

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
                                                     |||||   |||
Db        423 ----------------------------------------HisGlnLysGluLeu--- 427

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAAAT 852
                                                           |||:::   :::
Db        428 -------------------------------------------GluArgLeuAsp 431

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||   |||                  :::     :::             ||||||
Db        432 ArgGluArgIleArgAlaGluArgLeuGluLysGluArgLeuGluAlaGluArgLeuGlu 451

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAA 972
                  |||:::    :::    :::   |||::::::|||          ::::::
Db        452 ------LysGluArgLeuValIleLysLysLeuLysLysGluLysGluIleGlnLeuGlu 469

Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
              |||      |||:::::::::                  |||      |||
Db        470 AsnGluArgLeuGlnSerLeuLys--------------------GlnLysGlnLysGln 482

Qy       1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAA 1092
              |||:::|||          |||   ::::::                |||||||||||||:::
Db        483 GluGluGluGluGluGluGluGluGluGluGlyGluGluGluAspAspAsp 502

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
              |||
Db        503 Asp---------------------------------------------------------- 503

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT 1212
                        |||:::|||:::      |||:::   |||         :::::::::
Db        504 ------------GluIleGluGlu---GluGluGluGluGluGlnLysSerAsnValGlu 518

Qy       1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA------------AAAATAAAA 1260
              ||||||   |||:::      ||||||      |||:::            |||   :::
Db        519 GluGlyGluGluGluGluGluGluGluGluAspAspAspGluIleLysGlnGlu 538

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              :::   ::::::   |||:::   :::|||:::         |||   :::   |||:::
Db        539 GluGluGluGluGluGluGluGluAspSerIleIleThrProLysAsnSerSerIle 558

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
                 |||                              :::|||      :::   |||:::
Db        559 IleLysGlu--------------------------GluValThrSerAspIleGluGlu 569

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                                      :::         |||     :::         |||
Db        570 LeuGluThrAspThrSerPheGlnTyrSerSerAlaAsnSerSerMetMetThrAlaAsp 589

Qy       1441 GAGTTATCTGAAGATATAACTAAA 1464
              :::      |||:::      |||
Db        590 GlnAspGluGluGluGluAspLys 597
```

```
RESULT 26
Q6YA79_PLARE
ID   Q6YA79_PLARE PRELIMINARY;      PRT;  3427 AA.
AC   Q6YA79;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Reticulocyte binding-like protein 2b.
GN   Name=RBL2b;
OS   Plasmodium reichenowi.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5854;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=14698440; DOI=10.1016/j.molbiopara.2003.10.017;
RA   Rayner J.C., Huber C.S., Galinski M.R., Barnwell J.W.;
RT   "Rapid evolution of an erythrocyte invasion gene family: the
RT   Plasmodium reichenowi Reticulocyte Binding Like (RBL) genes.";
RL   Mol. Biochem. Parasitol. 133:287-296(2004).
DR   EMBL; AY166677; AAO38039.1; -; Genomic_DNA.
SQ   SEQUENCE   3427 AA;  404039 MW;  F7C66ED1296686C1 CRC64;

Alignment Scores:
Pred. No.:              1.03e-12      Length:          3427
Score:                  429.50        Matches:         140
Percent Similarity:     50.18%        Conservative:    136
Best Local Similarity:  25.45%        Mismatches:      171
Query Match:            17.28%        Indels:          103
DB:                     2             Gaps:            24

US-09-837-344-41 (1-1482) x Q6YA79_PLARE (1-3427)

Qy           1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
               ||||||:::              |||:::||||||:::   :::|||:::::::::::::::
Db        2672 GlnGluLysLeuLysAlaGlnGluGluGluArgLysGluArgGluArgIleGluLysGlu 2691

Qy          61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
               :::          |||::::::||||||   :::||||||   :::|||::::::
Db        2692 LysGln------GluLysGluArgLeuGluArgGluLysGlnGluGluGluArgLys--- 2708

Qy         121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------- 171
                 |||:::|||||||||   :::||||||   :::|||::::::   :::
Db        2709 ---GluArgGluArgLeuGluArgGluLysGlnGluGluGluArgLysGluLysGluArg 2727

Qy         172 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 231
               :::|||:::|||::::::   ||||||:::|||:::   ::::::          |||:::
Db        2728 IleGluLysGluLysGlnGluLysGluArgLeuGluArgGluLysGln------GluGlu 2745

Qy         232 GAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAAGATAGA 288
               ||||||:::   :::|||:::|||:::   |||:::|||         :::::::::
Db        2746 GluArgLysGluArgGluArgLeuGluArgGluLysGlnGluGlnGlnGlnLysGluGlu 2765

Qy         289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                  |||   :::   ||||||:::|||:::   |||:::::::|||          |||
Db        2766 GluLeuLysArgGlnGluGlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLys 2785

Qy         349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
                  :::   ||||||:::|||:::   |||:::::::|||          |||   :::
Db        2786 ArgGlnGluGlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGlu 2805

Qy         409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
               ||||||:::|||:::   |||:::::|||          |||   :::   ||||||:::
Db        2806 GlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLys 2825

Qy         469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
               |||:::   |||::::::|||          |||   :::   ||||||:::|||:::
Db        2826 GlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLysGlnAlaGln 2845

Qy         529 TTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
               |||::::::|||         :::|||   :::::::::   ||||||||||||
Db        2846 LeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLysGlnGluGlnGlnGlnLys 2865
```

```
Qy       580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db      2866 GluGluLeuLysArgGlnGluGlnGluLysGlnGluGlnGluGlnArgAsnIleGlnGlu 2885

Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
Db      2886 ----------------------LeuGluGluGlnLysLys---ProGluIleIleAsn 2896

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
Db      2897 GluAlaLeuValLysGlyAspLysIle-----------LeuGluGlySerAspGlnArg 2912

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
Db      2913 AsnMetGluLeu---SerLysProAsnIleSerMetAspAsnThrAspAsnSerProSer 2931

Qy       820 TCCAAGGAAATATCTATAATAGAAAAA---------ACAAATAGAGAATCTATTACAACA 870
Db      2932 SerAsnSer---GluIleIleGluSerAspAspIleAspAsnSerGluSerIleHisThr 2950

Qy       871 ------------AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAG 918
Db      2951 SerHisMetSerAspIleGluSerThrGlnThrAlaAspIleAsnAsnLeuGlnSerThr 2970

Qy       919 AAAGATGGTTCAATAAAACCAGAACAA----------------------------AAA 948
Db      2971 GlnThrValAspIleAsnAsnProGlnSerThrGlnThrAlaAspIleAsnAsnLeuGln 2990

Qy       949 GAAGATAAATCTGCTGACATACAAAAT--------------------------CATACA 981
Db      2991 SerThrGlnThrAlaAspIleAsnAsnLeuGlnSerThrGlnThrAlaAspIleAsnAsn 3010

Qy       982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT--- 1038
Db      3011 LeuGlnSerThrGlnThrSerAspIleAsnAsnLeuGlnSerThrGlnThrAlaAspIle 3030

Qy      1039 ------------------------------------------------GAAATAAGT 1047
Db      3031 AsnAsnLeuGlnSerThrGlnThrSerHisArgSerAsnThrHisGlyGlnGluIleSer 3050

Qy      1048 GCTGAATATGACGATTCATTA--------------ATAGATGAAGAAGAAGATGATGAA 1092
Db      3051 AspIleValAspAspGlnIleThrHisProSerAsnIleGlyLysGluLysIleThrHis 3070

Qy      1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
Db      3071 Asn---AspGluIleSerIleThrSerGluArgAsnAsnIleSerAsp------------ 3085

Qy      1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT 1212
Db      3086 --------------ValAsnAspHisSerGluSerSerAsnIlePheGluAsnGlyAsp 3100

Qy      1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCT------------GAAGAAAAAATAAAA 1260
Db      3101 SerThrIleAsnThrSerThrArgAspThrSerSerThrHisAspGluAlaHisIleSer 3120

Qy      1261 ---AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
Db      3121 ProSerGlyAsnAlaTyrAspArgValValSerAspAsnLysLysSerMetGlyGluAsn 3140

Qy      1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
Db      3141 IleLysAsp-----------LysLeuLysIleAspGluSerIleThrThrAspGluGln 3156

Qy      1378 AAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTG 1437
Db      3157 ------IleArg--------------LeuAspAspAsnSerAsnIleIleArgIleAsp 3169

Qy      1438 GATGAGTTATCTGAAGATATAACTAAATAT 1467
Db      3170 SerThrAspGlnArgAspValSerSerHis 3179
```

RESULT 27

```
Q54JG8_DICDI
ID   Q54JG8_DICDI PRELIMINARY;      PRT;  1082 AA.
AC   Q54JG8;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0187762;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000164; EAL63379.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1082 AA;  127407 MW;  9585384E168072C4 CRC64;
```

```
Alignment Scores:
Pred. No.:              1.53e-12      Length:         1082
Score:                  428.00        Matches:        142
Percent Similarity:     50.38%        Conservative:   122
Best Local Similarity:  27.10%        Mismatches:     158
Query Match:            17.22%        Indels:         102
DB:                     2             Gaps:           22

US-09-837-344-41 (1-1482) x Q54JG8_DICDI (1-1082)

Qy         4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG---------CAA 54
Db       583 GluArgGlnArgLysLeuGluGluGluLysLysGluGlnGluLysLysLysLeuGluGln 602

Qy        55 GAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
Db       603 GluArgGluAlaLysLeuGluAsnGluArgArgLeuArgGluGluLysLysGlnLysGlu 622

Qy       112 CAAAGCGATTTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
Db       623 ArgAlaAspArgGluGluLysGluArgLeuThrValGluArgLeuAlaLysGluArgGlu 642

Qy       169 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAG-----------------TTGCAAGAA 210
Db       643 LeuValGlnGlnGluArgGluLeuLysGluLysGlnLeuLeuAspLysIleLysLysAsp 662

Qy       211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
Db       663 LysGlnGluLysLeuGluLysGluLysLeuGluLysGluArgLeuGluAsnGluLys--- 681

Qy       271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT------ 324
```

```
Db        682 ---LeuGluAsnGluArgLeuGlyLysGluLysLeuGluLysGluLysLeuGluAsnGlu 700

Qy        325 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------ 375
                  ||||||   ||||||         |||:::|||:::   ::::::   :::
Db        701 ArgLeuGluAsnGluArgLeuGluAsnGluArgLeuGluAsnGluArgLeuGluAsnGlu 720

Qy        376 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------ 426
                  ||||||   ||||||      ||||||:::|||:::::::::::::   :::
Db        721 ArgLeuGluAsnGluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluAsnGly 740

Qy        427 ------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG---------------------- 459
                    ||||||:::|||:::|||   ||||||||||||
Db        741 GluLysLeuGluLysGluLysLeuGluLysGluLysLeuGluLysGluGluLeuGluAsn 760

Qy        460 ----------------------------CAAGAACAACAAAGCGATTTAGAACAAGAG 489
                                           |||:::||||||         ::::::|||
Db        761 LysArgAsnGluGluGluLeuLeuLeuLysGlnGlnGlnGlnGlnLysGluLysGluGlu 780

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
                  :::      |||||||||   ::::::::::::   :::   |||:::|||   :::
Db        781 LysGluGluLysGluLysGluGluLysGluLysGluGluLysGluLysGluGluLysGlu 800

Qy        550 AAAGAAAAGTTGCAA-----------------GAACAACAAAGCGATTTAGAACAAGAG 591
                  ||||||:::::::::                |||::::::   :::   |||:::
Db        801 LysGluArgIleGluLeuGluLysGlnLysAlaGluGluGluGluGluAlaGluLysLeu 820

Qy        592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTA---------------GAA 636
                  |||:::   ||||||:::   :::||||||:::   :::|||                |||
Db        821 ArgLysLeuLysGluGlnGluGluGluGlnLysLeuGluLeuLysArgIleArgAspGlu 840

Qy        637 CAAAGGAAGGCTGAT-----------------ACGAAAAAAAATTTAGAAAGAAAAAAG 678
                  |||:::|||   |||                  |||:::|||:::|||::::::|||
Db        841 GlnLysLysIleAspAspGluLysGluArgIleGluLysGluAsnIleGluLysGluLys 860

Qy        679 GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAA 738
                  :::      :::         ::::::         |||||||||            :::
Db        861 GlnArgIleGluAsnGlnGluLysGlu---------ArgLeuGlu-----------Lys 873

Qy        739 CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGAC 798
                  |||      |||   :::|||   :::                              |||
Db        874 LeuGlnGlnGluGlnGlnArgLeuPhe----------------------------Asp 883

Qy        799 AACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA---ATAGAAAAAACAAATAGA 855
                  :::         :::   ||||||||||||   |||   :::   |||      |||
Db        884 GluGln------GluLysLysSerLysGluIleGluIleGlnLeuLeuLysGluGluArg 901

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
                  |||   :::                        :::      |||      |||
Db        902 GluGluLeu----------------------GluAlaGluLysAlaArgLeuAlaSer 913

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
                     |||      :::               :::   :::   |||            :::
Db        914 LeuLysSerGluAlaAlaAlaAlaArgSerArgSerAsnSerSerThrSerAsnSerSer 933

Qy        976 CATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAG 1035
                  :::      :::   |||::::::      |||:::      :::
Db        934 AsnAsnSerSerSerSerAsnLeuAsnHis---AsnAsnThrAsnAsnAsnAsnAsnAsn 952

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGAC 1095
                  :::      :::         :::::::::      :::               ::::::
Db        953 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnThrSerThrAspAsn 972

Qy       1096 TTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGA 1155
                  |||:::|||            |||
Db        973 LeuAsnGluSerAsnGlySerValThrProProProThrSerValSerArgArgGlnSer 992

Qy       1156 ATTTATAAAGAACTAGAAGATTTGATAGAGAAA-------------------------- 1188
                  |||         :::      |||::::::|||
Db        993 IleSerSerGlyIleGlySerLeuLeuLysLysAlaValProSerSerProSerSerSer 1012

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                     |||         :::   |||:::   :::   |||||||||      |||::::::
Db       1013 LysGluGluLysLysGluLysAspAspThrLeuAsnLysSerSerAlaSerLeuThrLys 1032

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                  |||:::   |||:::   :::|||   ||||||                  |||
Db       1033 GluGluLysLysGluLysGluLysIleGluLys-----------GluLeuLysGluLys 1048

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                  :::|||          :::|||      ||||||   |||   :::|||         |||
Db       1049 GluLysArgGluLysGluGluArgLysLysLysGluLysValGluLys---------Lys 1065

Qy       1369 GAAAAGGAAAAA 1380
                  ||||||||||||
Db       1066 GluLysGluLys 1069

RESULT 28
Q54C75_DICDI
ID   Q54C75_DICDI PRELIMINARY;      PRT;  2205 AA.
AC   Q54C75;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0191797;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000274; EAL60912.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   2205 AA;  254185 MW;  681BA55EC6E7C909 CRC64;

Alignment Scores:
Pred. No.:              2.52e-12        Length:         2205
Score:                  422.00          Matches:        144
Percent Similarity:     51.00%          Conservative:   137
Best Local Similarity:  26.13%          Mismatches:     130
Query Match:            16.98%          Indels:         140
DB:                     2               Gaps:           27

US-09-837-344-41 (1-1482) x Q54C75_DICDI (1-2205)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
                ||||||||||      ||||||:::|||:::      ||||||:::|||::::::::::::
Db        610 LysGluLysLysLysLeuGluLysGluLysLeuGluLysGluArgLeuGluLysGluArg 629

Qy         64 AGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
                  :::   ||||||:::::::::|||   ||||||||||||::::::::::::
Db        630 LeuGluArgLeuGluLysGluLysLeuGluLysGluLysLeuGluLysGluLys------ 647
```

```
Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db       648 LeuGluLysGluLysLeuGluLysGluLysLeuGluLys-----------LeuGluLys 663

Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---TTAGAACAAGAGAGA 237
Db       664 GluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluArgLeuGluLysGluArg 683

Qy       238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
Db       684 LeuGluLysGluLysLeuGluLysGluArg------LeuGluLysGluArgLeuGluLys 701

Qy       298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db       702 GluArgLeuGluLysGluArg------LeuGluLysGluArgLeuGluLysGluArgLeu 719

Qy       358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db       720 GluArg-----------LeuGluLysGluArgLeuGluLysGluArgLeuGluLysGlu 735

Qy       418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
Db       736 Arg------LeuGluLysGluArgValGluLysGluArgLeuGluLysGluArgGln--- 752

Qy       478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------GAT 528
Db       753 ---GluLysGluArgLeuGluLysGluArgLeuGluLysGluLysSerLeuArgGluGln 771

Qy       529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       772 LeuGluLysGluArgLeuGluLysGluSerLeuGluLysGluArg------ValGluArg 789

Qy       589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db       790 GluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluLys 809

Qy       649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
Db       810 LeuLeuLysGluLysLeuGluLysGluLys-----------------------SerLeu 821

Qy       709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
Db       822 LysGluLysLeuGluLysGluLysIleGluLysAspLysGluLysGlu----------- 837

Qy       769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
Db       838 -----------------------AspAspAspGluAsnGluAsnGluAspLysMetAsp 849

Qy       829 ATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
Db       850 ThrAsp------GlyLysValAsnArgSerSerLysThr---------GlnGluGlnArg 864

Qy       889 GATATA--------------CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db       865 AspLeuGluAsnLeuLysSerArgGluGlnArgLysArgGluLysGluGluAspLeuMet 884

Qy       934 AAACCAGAACAAAAAGAAGATAAATCT-----------------------GCTGACATA 969
Db       885 TyrTyrGluGlnAsnLysAsnLysSerAspGlnAsnArgGlyGlyGlyGlyGlyLeu 904

Qy       970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
Db       905 ValSerGlnPheLeuGlnGlnSerAsnProLeuSerGlnLeuGlnLeuGlnSerAsnLys 924

Qy      1030 TATGAGGAT---GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
Db       925 IleValAspLeuGluValIleGluAspTyr---------IleAlaGluArgLysArgGlu 941

Qy      1087 ------GATGAAGACTTAGACGAATTTAAGCCTATTGTG-------------------- 1119
Db       942 TrpPheGluLysAsnIleAspSerLeuLysLysMetAlaLeuProLeuTrpAsnLeuSer 961

Qy      1119 ------------------------------------------------------------ 1119

Db       962 ArgLysAsnMetGluValIleLysIleArgGluLysArgTyrPheGluTyrAsnAlaArg 981

Qy      1120 ---------CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTA 1170
Db       982 LysCysLysLeuPheAsnAspPheGlnTyrGluPheSerProLysValTyrLysAlaIle 1001

Qy      1171 GAAGATTTGATAGAG---------AAAAATGAAAATTTAGATGATTTA------------ 1209
Db      1002 SerGlnThrPheGluAspPheTrpLysGluGluTrpIleValAspMetLeuHisSerGln 1021

Qy      1210 ---------GATGAAGGAATAGAAAAATCATCA------------------------GAA 1236
Db      1022 LysProProProGluTyrValGluSerProSerGlnIleLysGlnGlnGlnGlnGlnGln 1041

Qy      1237 GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG--------- 1287
Db      1042 GlnGlnGlnGlnGlnGlnGlnGlnSerSerSerAlaGlnLysSerLysLeuIleVal 1061

Qy      1288 --------------------GATAATAATTTTAAACCAAATGAT--------------- 1311
Db      1062 ValGlnGlnGluThrAspAspAspAspAspAspSerProAspAspTrpAsnGluAspAsn 1081

Qy      1312 --------------AAAAGTTTGTATGAT---------GAGCATATTAAAAAATATAAA 1347
Db      1082 LeuLeuSerGlnGlnSerSerLeuHisGluAsnLysGlyGluAsnGluAsnAspSerGlu 1101

Qy      1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db      1102 AsnGluSerGluSerGluSerGluLysGlnLys 1112
```

RESULT 29
Q50RA3_ENTHI
ID Q50RA3_ENTHI PRELIMINARY; PRT; 762 AA.
AC Q50RA3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=295.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabbinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000803; EAL44122.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 762 AA; 94161 MW; A406F044AB7E1F84 CRC64;

Alignment Scores:

Pred. No.:	4.34e-12	Length:	762
Score:	419.00	Matches:	130
Percent Similarity:	48.95%	Conservative:	103
Best Local Similarity:	27.31%	Mismatches:	149
Query Match:	16.85%	Indels:	94

```
DB:                 2            Gaps:          21

US-09-837-344-41 (1-1482) x Q50RA3_ENTHI (1-762)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAAGTTGCAAGAACAA 60
              :::   |||:::   ::::::::::::::          |||   ||||||:::|||
Db        109 LysSerGlnGluGluGluValGlnLysGlnGluGluGluLysArgLysLeuGluGluIle 128

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||         |||:::            |||   ||||||:::|||::::::
Db        129 GlnArgGlnGluGluGlu-----------LysArgLysLeuGluGluGluGluGluLys 144

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---GATCTAGAA 177
              ::::::   |||          |||   :::|||:::|||::::::          ||||||
Db        145 MetLysLeuGluGluGluGluLysAsnArgLeuGluGluGluGluLysArgLysLeuGlu 164

Qy        178 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
              :::|||   ::::::   ||||||:::::::|||   |||:::   ::::::::::::::
Db        165 GluGluIleGlnLysGlnLysGluGluIleGlnArgGlnGluGluGluIleGlnLysGln 184

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 294
                  |||   ||||||:::|||:::::::::   ||||||
Db        185 GluGluGluLysMetLysLeuGluGluGluLysAsnArgLeuGlu--------------- 199

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 354
                 |||::::::|||:::|||:::   :::   :::::::::          |||   |||
Db        200 ---GluGluIleGlnLysGlnGluGluGluArgGlnLysGlnGluGluGluLysAsnLys 218

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
              |||:::|||:::              |||:::|||:::|||         ||||||:::|||
Db        219 LeuGluGluGluIleGlnLysGlnGluGluGluLysArg---------LysLeuGluGlu 235

Qy        415 CAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              ::::::      ||||||:::   :::      ||||||::::::|||:::|||:::
Db        236 GluLysLysArgLeuGluGluThrLysArgGlnLysGluGluIleGlnLysGlnGluGlu 255

Qy        475 GATTTAGAACAAGAG------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
                :::||||||      :::|||   ||||||::::::|||   |||:::   :::
Db        256 ArgGlnLysGlnGluGluIleGlnArgGlnLysGluGluIleGlnArgGlnGluGluGlu 275

Qy        529 ------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
                   ||||||          |||   |||   :::|||   |||
Db        276 GluArgLysArgGlnGluGluGluArgLysArgLysGluGluGluIleGlnArgGlnLys 295

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
              |||:::||||||:::   :::|||::::::|||   |||:::          |||:::|||
Db        296 GluGluGluArgLysLysGlnGluGluIleGlnArgGlnLys---------GluGluArg 312

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
              |||          |||:::      |||||||||::::::
Db        313 Lys------ArgLysGluGluGluGluArgLysArgGln--------------------- 323

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
                                                      |||   ||||||.
Db        324 ---------------------------------------------GluGluGluArg--- 327

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC 822
                                        |||                  |||
Db        328 ----------------------------GlnLysGlnGluGluIleGlnArgGlnLys 337

Qy        823 AAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGA 882
              :::|||          |||::::::          |||   |||              |||
Db        338 GluGlu---------IleGlnArgGlnLysGluGluGluIleGlnArgGlnLysGluGlu 354

Qy        883 CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAA 942
              |||:::        |||          |||||||||:::      :::|||   :::
Db        355 ArgLys------ArgLysGluGluGluArgGluLysLysGluArgGluLeuLysArgLys 372

Qy        943 CAAAAAGAAGATAAATCTGCT---------GACATACAAAATCATACATTAGAGACAGTA 993
                 :::||||||:::                |||   :::|||          |||   :::
Db        373 LeuGluGluAspArgLysValLeuGluGluAspLysGluAsnGlnLysAspGluLysIle 392

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
                  |||   |||            :::|||:::      |||||||||:::
Db        393 IleIleLysAsp-----------GluIleAsnIleLeuGluAspGluVal--------- 405

Qy       1054 TATGACGATTCATTAATAGATGAA--------------GAAGAAGATGATGAAGACTTA 1098
                          |||:::::::|||                |||||:::   :::   :::
Db        406 ---------HisLeuLeuGluGluLysLysGlyGluLeuGluGluGluLysGlnArgIle 422

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
              ||||||                ::::          |||:::::|||:::
Db        423 AspGlu--------------GluMetHisLysPheGluGluGluLys------------ 433

Qy       1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGA 1218
                 |||   :::|||:::   |||
Db        434 ---LysArgIleGluGluGluIleAsnAsnLysArgGlyAlaSerAlaSerArgMetTyr 452

Qy       1219 ATAGAAAAATCATCAGAA---GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
                    |||      |||   ||||||   :::   ||||||:::      :::|||
Db        453 PheAlaLysLysGlnGluIleGluLeuGlnGlnAsnLysIleGluAsnLysGlnLysGlu 472

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATGAT---------AAAAGTTTGTATGAT 1326
                    |||                  |||   |||                |||      :::
Db        473 AlaThrThrAsnLysMetLysIleLysLysAsnGlnSerGluIleLysArgLysGlnGlu 492

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
              |||   |||||||||   |||      |||:::::::::|||   |||
Db        493 GluIleIleLysLysGluLysGluIleLysGluIleAspLysArgLys 508

RESULT 30
Q54MJ1_DICDI
ID   Q54MJ1_DICDI PRELIMINARY;      PRT;  1061 AA.
AC   Q54MJ1;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0186744;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
CC   -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
DR   EMBL; AAFI01000127; EAL64490.1; -; Genomic_DNA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
KW   Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ   SEQUENCE   1061 AA;  128819 MW;  4B2361B2D1531D0D CRC64;

Alignment Scores:
Pred. No.:              4.05e-12       Length:        1061
```

Score:	419.00	Matches:	146
Percent Similarity:	49.58%	Conservative:	147
Best Local Similarity:	24.70%	Mismatches:	178
Query Match:	16.85%	Indels:	120
DB:	2	Gaps:	25

US-09-837-344-41 (1-1482) x Q54MJ1_DICDI (1-1061)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA--- 57
              :::|||:::|||         :::||||||:::      :::||||||   :::|||
Db        266 LysGluGluGlnLeuLysArgAspGlnGluLysLeuGluArgGluLysLysArgGluTyr 285

Qy         58 ------------------------------------CAACAAAGCGATTTAGAACAA 78
                                                  ||||||      ||||||:::
Db        286 HisGlnGlnGlnProGlnGlnThrGlnIleGlnIleHisGlnGlnLysMetLeuGluLys 305

Qy         79 GATAGACTTGCTAAAGAAAAG------------TTACAAGAGCAGCAAAGCGATTTA--- 123
              :::|||||||||:::|||:::            |||   |||:::::::      |||
Db        306 GluArgLeuAlaGlnGluGluLysGluArgLeuLeuValGluGluArgGluArgLeuAsn 325

Qy        124 --------------------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 156
                                         |||:::|||||||||   :::|||:::   :::
Db        326 AlaThrArgValGluArgGluLysGlnGluLysGluArgLeuAspArgGluArgLysGlu 345

Qy        157 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAA 213
              :::|||:::   ||||||:::||||||         |||:::|||:::   |||:::
Db        346 ArgGluGlnAlaArgLeuGluLysGluArgLeuGluSerGluArgLeuGluLysGluArg 365

Qy        214 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273
              |||:::   :::|||:::||||||:::      |||:::::::::::::::::   |||
Db        366 GlnAlaArgValGluArgGluArgLysGluPheGluArgIleGluLysGluLysArgAsp 385

Qy        274 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 333
                ||||||   ||||||     |||:::|||:::   ::::::      |||::::::
Db        386 ArgGluGlnThrArgLeuGluAsnGluArgLeuGluAsnGluArg------LeuLysLys 403

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGA 390
              |||:::      ||||||:::|||:::   ::::::   :::   |||   |||||||||
Db        404 GluLysLeuGluLysGluArgLeuGluAsnGluArgPheGluLysGluArgGlnGluArg 423

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGACTTGCT 447
                  |||:::|||:::   ::::::   :::   |||   :::|||||||||
Db        424 IleGluAsnGluArgLeuGluAsnGluArgPheGluArgGluArgLysGluArgLeuGlu 443

Qy        448 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 507
              |||||||||:::::::::::::::   :::      |||   |||   ||||||||||||
Db        444 LysGluLysIleGluLysGluGluArgGlu---AsnGlnLeuArgLeuAlaLysGluLys 462

Qy        508 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 567
                 :::|||:::::                |||||||||:::   ||||||:::|||   |||
Db        463 ---GluGluLysGluGluLysGluArgGlnGluArgGlnLeuLysGluArgLeu---Glu 480

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
              ::::::      ||||||:::||||||:::   :::|||:::   :::|||:::   |||
Db        481 ArgLysGluLysLeuGluLysGluArgLysGluArgGluGluLysGluGluLysAspArg 500

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA-------------- 672
                  |||   :::        |||:::   |||:::|||
Db        501 ---------GlnGluArgLeuGlnLeuLysGluArgLeuAspArgGluArgLysGluArg 517

Qy        673 ---AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ::||||||      :::         :::|||         |||   |||
Db        518 LeuGluLysGluArgLysGluArgGluGluLysAspGluLysAspArgGlnGluArgLeu 537

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA--------- 780
              |||:::         |||   ||||||        :::      :::
Db        538 GlnIleLysGluArgLeuGluArgGluArgLysGluArgLeuGluLysGluArgLeuGlu 557

Qy        781 ---TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATA 837
                  |||         :::        ::::::   ||||||
Db        558 TyrGluArgLeuGluThrLeuArgLysGluLysGluLysGluArgLysGluLysPheGlu 577

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                    :::      ||||||   :::         ::::::   :::   :::
Db        578 ThrAlaArgLeuGluArgGluLysLeuGluLeuGluLeuLysGluLysAspGluLysLys 597

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::         |||||||||:::         |||       |||   ::::::   :::
Db        598 Arg---------GluGluLysArgLeuAlaLysIleIleGluGluAsnArgLysLysGlu 614

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                ||| :::                 |||         :::|||      ::::::
Db        615 LysAlaLeuValAspLysLysLeuGluGluGlnGluArgLeuSer------HisGluLeu 632

Qy       1018 CAAATAAGTAAGTATGAG--------------GATGAAATAAGTGCTGAATATGACGAT 1062
              :::|||   |||   :::                :::::::::      :::   :::
Db        633 LysIleGlnLysGluLysGluIleLeuLysMetGluLysLeuAspLeuGlnCysGluLys 652

Qy       1063 TCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC--------------------- 1101
                   :::|||:::|||:::   :::      ::::::
Db        653 LysLysLeuAspLysGluLysLysGluLeuGlnIleGluLysGluLeuSerGlnLysGln 672

Qy       1102 ---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
                 :::|||      :::   |||   :::      |||   ::::::   |||
Db        673 LeuGlnPheIleGluLeuLysGlnGlnGluIleAsnGlnGlnGlnGlnGluIleAsnHis 692

Qy       1159 TATAAAGAACTAGAAGAT--------------------TTGATAGAGAAAAATGAAAAT 1197
                 ::::::   :::                        |||:::   |||
Db        693 GlnGlnGlnGlnGlnGlnGlnThrGlnGlnGlnIleProLeuLeuThrLysGlnProThr 712

Qy       1198 TTAGATGATTTA---------------------------------------------- 1209
                 ::::::::::::
Db        713 ValGluGluIleGluTyrGluArgValGlnGluGluIleLysGlnLysLysLeuGluArg 732

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
              ::::::   :::::::::      ||||||      :::|||:::::|||:::   :::
Db        733 GluGlnArgLeuLysGlnGlnLeuGluGluGluGluLysGluArgLeuLysArgGlnGlu 752

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
              :::   :::|||   |||:::                 ||||||   |||   :::|||
Db        753 GluArgGlnLysArgLysGluGln------------ArgAspLysGluLeuLysGluGlu 768

Qy       1330 CATATT--------------------------AAAAAATATAAAAATGATAAG------ 1356
                 :::                         ||||||   |||   ::::::
Db        769 GluLeuArgIleGluAlaGluArgGluThrArgLysLysLeuLysGluGluArgGluLeu 788

Qy       1357 CAGGTTAATAAGGAA-----------------------------------AAGGAAAAA 1380
              |||      ||||||                                   |||||||||
Db        789 GlnPheLysLysGluGlnValGlnLeuLeuLeuGluArgAspLysArgValLysGluLys 808

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGAC---GGAGACAATGAAATTTTACAGATCGTG 1437
                 :::|||         :::   :::        |||   |||:::      ||||||
Db        809 GlnLeuLysAspGlnLysTyrThrTyrLysIleAsnAspThrGluLeuGluHisIleVal 828

Qy       1438 GATGAGTTATCTGAAGATATAACTAAATATTTT 1470
              ::::::      :::|||:::||||||   |||
Db        829 GluAsp------LysAspLeuThrLysGluPhe 837
```

Search completed: December 21, 2005, 11:26:19
Job time : 424.484 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 2.0261 Seconds
(without alignments)
5204.614 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4883371

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22895/app_query.fasta_1.3740
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_821_@runat_21122005_093215_22895 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 10:45:00
Job time : 2.0261 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 0.415663 Seconds
(without alignments)
4773.622 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1143330

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093212_22815/app_query.fasta_1.3740
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_161_@runat_21122005_093212_22815 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 09:37:12
Job time : 0.415663 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:23 ; Search time 1.89759 Seconds
(without alignments)
5284.546 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3732626

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22847/app_query.fasta_1.3740
-DB=Published_Applications_AA_Main -QFMT=fastan -SUFFIX=rapbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_781_@runat_21122005_093214_22847
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 11:00:44
Job time : 1.89759 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:38:31 ; Search time 0.0682731 Seconds
(without alignments)
2507.145 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 53982 seqs, 7132107 residues

Total number of hits satisfying chosen parameters: 107923

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22867/app_query.fasta_1.3740
-DB=Published_Applications_AA_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_2_@runat_21122005_093214_22867
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	20	95.2	9	7	US-11-097-864-11	Sequence 11, Appl
c 2	20	95.2	9	7	US-11-097-864-150	Sequence 150, App
c 3	20	95.2	9	7	US-11-097-864-353	Sequence 353, App
c 4	20	95.2	9	7	US-11-097-864-463	Sequence 463, App
c 5	20	95.2	9	7	US-11-097-864-587	Sequence 587, App
c 6	20	95.2	9	7	US-11-097-912-11	Sequence 11, Appl
c 7	20	95.2	9	7	US-11-097-912-150	Sequence 150, App
c 8	20	95.2	9	7	US-11-097-912-353	Sequence 353, App
c 9	20	95.2	9	7	US-11-097-912-463	Sequence 463, App
c 10	20	95.2	9	7	US-11-097-912-587	Sequence 587, App
c 11	20	95.2	10	7	US-11-097-864-337	Sequence 337, App
c 12	20	95.2	10	7	US-11-097-864-509	Sequence 509, App
c 13	20	95.2	10	7	US-11-097-912-337	Sequence 337, App
c 14	20	95.2	10	7	US-11-097-912-509	Sequence 509, App
c 15	20	95.2	46	6	US-10-467-657-8088	Sequence 8088, Ap
16	20	100.0	71	7	US-11-132-142-17	Sequence 17, Appl
17	20	100.0	115	7	US-11-156-909-11	Sequence 11, Appl
18	20	100.0	133	7	US-11-156-909-10	Sequence 10, Appl
c 19	20	95.2	135	6	US-10-467-657-2028	Sequence 2028, Ap
c 20	20	95.2	136	6	US-10-793-626-912	Sequence 912, App
21	20	100.0	149	7	US-11-156-909-31	Sequence 31, Appl
c 22	20	95.2	166	6	US-10-467-657-3512	Sequence 3512, Ap
23	20	100.0	169	6	US-10-793-626-1534	Sequence 1534, Ap
24	20	100.0	173	7	US-11-156-909-29	Sequence 29, Appl
c 25	20	95.2	176	6	US-10-793-626-362	Sequence 362, App
26	20	100.0	184	6	US-10-467-657-5748	Sequence 5748, Ap
c 27	20	95.2	208	7	US-11-055-822-462	Sequence 462, App
c 28	20	95.2	208	7	US-11-055-822-704	Sequence 704, App
29	20	100.0	212	6	US-10-821-234-1430	Sequence 1430, Ap
c 30	20	95.2	223	6	US-10-878-556A-167	Sequence 167, App
31	20	100.0	230	6	US-10-793-626-3242	Sequence 3242, Ap
32	20	100.0	242	7	US-11-100-183-12	Sequence 12, Appl
33	20	100.0	249	6	US-10-467-657-2256	Sequence 2256, Ap
34	20	100.0	259	7	US-11-000-463-276	Sequence 276, App
c 35	20	95.2	266	6	US-10-467-657-5258	Sequence 5258, Ap
36	20	100.0	285	7	US-11-179-843-3	Sequence 3, Appli
c 37	20	95.2	287	6	US-10-467-657-2960	Sequence 2960, Ap
38	20	100.0	299	6	US-10-131-826A-366	Sequence 366, App
39	20	100.0	299	7	US-11-000-463-275	Sequence 275, App
40	20	100.0	299	7	US-11-000-463-747	Sequence 747, App
41	20	100.0	299	7	US-11-000-463-748	Sequence 748, App
42	20	100.0	299	7	US-11-173-037-7	Sequence 7, Appli
c 43	20	95.2	299	7	US-11-112-882-30	Sequence 30, Appl
44	20	100.0	300	7	US-11-025-834A-21	Sequence 21, Appl
45	20	100.0	316	6	US-10-793-626-1350	Sequence 1350, Ap

ALIGNMENTS

RESULT 1
US-11-097-864-11
; Sequence 11, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006205
; CURRENT APPLICATION NUMBER: US/11/097,864
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-864-11

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0

```
DB:                    7              Gaps:          0

US-09-837-344-42 (1-12) x US-11-097-864-11 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          4 ValTyrAsnLeu 7

RESULT 2
US-11-097-864-150
; Sequence 150, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-150

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-42 (1-12) x US-11-097-864-150 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          6 ValTyrAsnLeu 9

RESULT 3
US-11-097-864-353
; Sequence 353, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 353
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-353

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-42 (1-12) x US-11-097-864-353 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 4
US-11-097-864-463
; Sequence 463, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 463
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-463

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-42 (1-12) x US-11-097-864-463 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5

RESULT 5
US-11-097-864-587
; Sequence 587, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
```

```
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 587
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-587

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-864-587 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          5 ValTyrAsnLeu 8

RESULT 6
US-11-097-912-11
; Sequence 11, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-11

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-11 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          4 ValTyrAsnLeu 7

RESULT 7
US-11-097-912-150
; Sequence 150, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-150

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-150 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          6 ValTyrAsnLeu 9

RESULT 8
US-11-097-912-353
; Sequence 353, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 353
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-353

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
```

```
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-353 (1-9)

Qy          12 ATATATAACCTT 1
               :::||||||||||
Db           1 ValTyrAsnLeu 4

RESULT 9
US-11-097-912-463
; Sequence 463, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 463
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-463

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-463 (1-9)

Qy          12 ATATATAACCTT 1
               :::||||||||||
Db           2 ValTyrAsnLeu 5

RESULT 10
US-11-097-912-587
; Sequence 587, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 587
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-587

Alignment Scores:
Pred. No.:              1.56e+06        Length:         9
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-587 (1-9)

Qy          12 ATATATAACCTT 1
               :::||||||||||
Db           5 ValTyrAsnLeu 8

RESULT 11
US-11-097-864-337
; Sequence 337, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 337
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-337

Alignment Scores:
Pred. No.:              200             Length:         10
Score:                  20.00           Matches:        3
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  75.00%          Mismatches:     0
Query Match:            95.24%          Indels:         0
DB:                     7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-864-337 (1-10)

Qy          12 ATATATAACCTT 1
               :::||||||||||
Db           1 ValTyrAsnLeu 4

RESULT 12
US-11-097-864-509
; Sequence 509, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
```

```
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-509

Alignment Scores:
Pred. No.:              200          Length:         10
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
DB:                     7            Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-864-509 (1-10)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5

RESULT 13
US-11-097-912-337
; Sequence 337, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 337
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-337

Alignment Scores:
Pred. No.:              200          Length:         10
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
DB:                     7            Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-337 (1-10)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 14
US-11-097-912-509
; Sequence 509, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-509

Alignment Scores:
Pred. No.:              200          Length:         10
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
DB:                     7            Gaps:           0

US-09-837-344-42 (1-12) x US-11-097-912-509 (1-10)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5

RESULT 15
US-10-467-657-8088
; Sequence 8088, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8088
;   LENGTH: 46
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8088

Alignment Scores:
Pred. No.:              250          Length:         46
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
```

```
DB:                    6             Gaps:          0

US-09-837-344-42 (1-12) x US-10-467-657-8088 (1-46)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db         11 ValTyrAsnLeu 14

RESULT 16
US-11-132-142-17
; Sequence 17, Application US/11132142
; Publication No. US20050276818A1
; GENERAL INFORMATION:
;  APPLICANT: The Burnham Institute
;  APPLICANT:  Sikora, Sergey
;  APPLICANT:  Godzik, Adam
;  TITLE OF INVENTION: UNCHARACTERIZED ORF3 IN SARS-CORONAVIRUS IS A
;  TITLE OF INVENTION:  CYCLIC-AMP-DEPENDENT KINASE AND A TARGET FOR SARS THERAPY
;  FILE REFERENCE: 8014-011-US
;  CURRENT APPLICATION NUMBER: US/11/132,142
;  CURRENT FILING DATE:  2005-05-17
;  PRIOR APPLICATION NUMBER: 60/571,698
;  PRIOR FILING DATE: 2004-05-17
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 17
;   LENGTH: 71
;   TYPE: PRT
;   ORGANISM: Arabidopsis sp.
US-11-132-142-17

Alignment Scores:
Pred. No.:             267           Length:        71
Score:                 20.00         Matches:       4
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           100.00%       Indels:        0
DB:                    7             Gaps:          0

US-09-837-344-42 (1-12) x US-11-132-142-17 (1-71)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         36 LysValIleTyr 39

RESULT 17
US-11-156-909-11
; Sequence 11, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 11
;   LENGTH: 115
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Mature Protein for Clone 3.4
US-11-156-909-11

Alignment Scores:
Pred. No.:             286           Length:        115
Score:                 20.00         Matches:       4
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           100.00%       Indels:        0
DB:                    7             Gaps:          0

US-09-837-344-42 (1-12) x US-11-156-909-11 (1-115)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        102 LysValIleTyr 105

RESULT 18
US-11-156-909-10
; Sequence 10, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 10
;   LENGTH: 133
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Precursor Protein for Clone 3.4
US-11-156-909-10

Alignment Scores:
Pred. No.:             292           Length:        133
Score:                 20.00         Matches:       4
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           100.00%       Indels:        0
DB:                    7             Gaps:          0

US-09-837-344-42 (1-12) x US-11-156-909-10 (1-133)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        120 LysValIleTyr 123

RESULT 19
US-10-467-657-2028
; Sequence 2028, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2028
;   LENGTH: 135
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2028

Alignment Scores:
Pred. No.:             293           Length:        135
Score:                 20.00         Matches:       3
```

```
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  75.00%          Mismatches:    0
Query Match:            95.24%          Indels:        0
DB:                     6               Gaps:          0

US-09-837-344-42.(1-12) x US-10-467-657-2028 (1-135)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db         85 ValTyrAsnLeu 88

RESULT 20
US-10-793-626-912
; Sequence 912, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 912
;   LENGTH: 136
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-912

Alignment Scores:
Pred. No.:              293             Length:        136
Score:                  20.00           Matches:       3
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  75.00%          Mismatches:    0
Query Match:            95.24%          Indels:        0
DB:                     6               Gaps:          0

US-09-837-344-42 (1-12) x US-10-793-626-912 (1-136)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db        123 ValTyrAsnLeu 126

RESULT 21
US-11-156-909-31
; Sequence 31, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 31
;   LENGTH: 149
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Mature Protein with His-tag, Clone 3.4
US-11-156-909-31

Alignment Scores:
Pred. No.:              297             Length:        149
Score:                  20.00           Matches:       4
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     7               Gaps:          0

US-09-837-344-42 (1-12) x US-11-156-909-31 (1-149)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        136 LysValIleTyr 139

RESULT 22
US-10-467-657-3512
; Sequence 3512, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3512
;   LENGTH: 166
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3512

Alignment Scores:
Pred. No.:              302             Length:        166
Score:                  20.00           Matches:       3
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  75.00%          Mismatches:    0
Query Match:            95.24%          Indels:        0
DB:                     6               Gaps:          0

US-09-837-344-42 (1-12) x US-10-467-657-3512 (1-166)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 23
US-10-793-626-1534
; Sequence 1534, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1534
;   LENGTH: 169
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-1534
```

```
Alignment Scores:
Pred. No.:              303          Length:        169
Score:                  20.00        Matches:       4
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-837-344-42 (1-12) x US-10-793-626-1534 (1-169)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         57 LysValIleTyr 60

RESULT 24
US-11-156-909-29
; Sequence 29, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 29
;   LENGTH: 173
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Precursor protein with His-tag, Clone 3.4
US-11-156-909-29

Alignment Scores:
Pred. No.:              304          Length:        173
Score:                  20.00        Matches:       4
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     7            Gaps:          0

US-09-837-344-42 (1-12) x US-11-156-909-29 (1-173)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        160 LysValIleTyr 163

RESULT 25
US-10-793-626-362
; Sequence 362, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 362
;   LENGTH: 176
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-362

Alignment Scores:
Pred. No.:              305          Length:        176
Score:                  20.00        Matches:       3
Percent Similarity:     100.00%      Conservative:  1
Best Local Similarity:  75.00%       Mismatches:    0
Query Match:            95.24%       Indels:        0
DB:                     6            Gaps:          0

US-09-837-344-42 (1-12) x US-10-793-626-362 (1-176)

Qy         12 ATATATAACCTT 1
                :::||||||||||
Db          1 ValTyrAsnLeu 4

RESULT 26
US-10-467-657-5748
; Sequence 5748, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5748
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5748

Alignment Scores:
Pred. No.:              307          Length:        184
Score:                  20.00        Matches:       4
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-837-344-42 (1-12) x US-10-467-657-5748 (1-184)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        177 LysValIleTyr 180

RESULT 27
US-11-055-822-462
; Sequence 462, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
```

```
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 462
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-462

Alignment Scores:
Pred. No.:                312             Length:         208
Score:                    20.00           Matches:        3
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    75.00%          Mismatches:     0
Query Match:              95.24%          Indels:         0
DB:                       7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-055-822-462 (1-208)

Qy          12 ATATATAACCTT 1
               :::|||||||||
Db          14 ValTyrAsnLeu 17

RESULT 28
US-11-055-822-704
; Sequence 704, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 704
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-704

Alignment Scores:
Pred. No.:                312             Length:         208
Score:                    20.00           Matches:        3
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    75.00%          Mismatches:     0
Query Match:              95.24%          Indels:         0
DB:                       7               Gaps:           0

US-09-837-344-42 (1-12) x US-11-055-822-704 (1-208)

Qy          12 ATATATAACCTT 1
               :::|||||||||
Db          14 ValTyrAsnLeu 17

RESULT 29
US-10-821-234-1430
; Sequence 1430, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1430
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1430

Alignment Scores:
Pred. No.:                313             Length:         212
Score:                    20.00           Matches:        4
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-837-344-42 (1-12) x US-10-821-234-1430 (1-212)

Qy           1 AAGGTTATATAT 12
               ||||||||||||
Db          64 LysValIleTyr 67

RESULT 30
US-10-878-556A-167
; Sequence 167, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 167
;   LENGTH: 223
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_hum/kad4_human
;   DATABASE ENTRY DATE: 1992-08-01
US-10-878-556A-167

Alignment Scores:
Pred. No.:              315           Length:        223
Score:                  20.00         Matches:       3
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  75.00%        Mismatches:    0
Query Match:            95.24%        Indels:        0
DB:                     6             Gaps:          0

US-09-837-344-42 (1-12) x US-10-878-556A-167 (1-223)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db        135 ValTyrAsnLeu 138


Search completed: December 21, 2005, 11:01:55
Job time : 1.06827 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:27:59 ; Search time 69.5382 Seconds
(without alignments)
9809.312 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	12	100.0	12	6	AR162971	AR162971 Sequence
2	12	100.0	12	6	AR162975	AR162975 Sequence
3	12	100.0	12	6	AR178260	AR178260 Sequence
4	12	100.0	12	6	AR178264	AR178264 Sequence
5	12	100.0	23	6	A36617	A36617 Sequence 26
6	12	100.0	104	10	G32838	G32838 A009U15 Hum
c 7	12	100.0	113	10	CR378386	CR378386 Arabidops
8	12	100.0	139	10	BX294638	BX294638 Arabidops
c 9	12	100.0	146	6	BD043274	BD043274 Sequence
c 10	12	100.0	146	6	AX907741	AX907741 Sequence
c 11	12	100.0	169	2	AF487987	AF487987 Triatoma
c 12	12	100.0	184	6	BD268181	BD268181 Gene to b
c 13	12	100.0	184	6	AR210010	AR210010 Sequence
14	12	100.0	188	6	AX196591	AX196591 Sequence
15	12	100.0	190	10	G23922	G23922 human STS W
c 16	12	100.0	200	10	AB153981	AB153981 Homo sapi
17	12	100.0	201	15	ATH531582	AJ531582 Arabidops
18	12	100.0	201	15	ATH531598	AJ531598 Arabidops
19	12	100.0	225	10	G31833	G31833 SpAGH1 Norw
c 20	12	100.0	226	15	AY020751	AY020751 Oryza sat
21	12	100.0	227	9	MMU409475	AJ409475 Mus muscu
c 22	12	100.0	232	10	AB143702	AB143702 Homo sapi
23	12	100.0	232	15	AY020956	AY020956 Oryza sat
24	12	100.0	234	13	AY114172	AY114172 Human her
c 25	12	100.0	236	10	G37836	G37836 wh_hsp1 Pla
c 26	12	100.0	236	10	G44501	G44501 C9M106 Plas
c 27	12	100.0	238	10	AU026951	AU026951 Rattus no
c 28	12	100.0	242	6	BD028743	BD028743 Sequence
c 29	12	100.0	242	6	AX893210	AX893210 Sequence
30	12	100.0	255	6	CQ457358	CQ457358 Sequence
31	12	100.0	255	10	G11552	G11552 human STS S
32	12	100.0	255	10	G26557	G26557 human STS S
c 33	12	100.0	264	10	BV344514	BV344514 S230P6149
34	12	100.0	267	6	AX311604	AX311604 Sequence
c 35	12	100.0	277	6	CQ475242	CQ475242 Sequence
c 36	12	100.0	281	10	AB152907	AB152907 Homo sapi
37	12	100.0	288	6	CQ748906	CQ748906 Sequence
c 38	12	100.0	289	6	CQ457286	CQ457286 Sequence
c 39	12	100.0	293	4	AF080181	AF080181 Procyon l
c 40	12	100.0	294	6	BD042247	BD042247 Sequence
c 41	12	100.0	294	6	AX906714	AX906714 Sequence
c 42	12	100.0	295	4	AF080174	AF080174 Procyon l
c 43	12	100.0	295	4	AF080175	AF080175 Procyon l
c 44	12	100.0	295	4	AF080178	AF080178 Procyon l
c 45	12	100.0	295	4	AF080180	AF080180 Procyon l

ALIGNMENTS

RESULT 1
AR162971
LOCUS AR162971 12 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 42 from patent US 6270771.
ACCESSION AR162971
VERSION AR162971.1 GI:16233436
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 12)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6270771-A 42 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .12
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.3e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 1 AAGGTTATATAT 12

RESULT 2
AR162975
LOCUS AR162975 12 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 46 from patent US 6270771.
ACCESSION AR162975
VERSION AR162975.1 GI:16233442
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 12)

```
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 46 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
AR178260
LOCUS       AR178260                  12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 42 from patent US 6319502.
ACCESSION   AR178260
VERSION     AR178260.1  GI:20219398
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 42 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
AR178264
LOCUS       AR178264                  12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 46 from patent US 6319502.
ACCESSION   AR178264
VERSION     AR178264.1  GI:20219402
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 46 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
A36617
LOCUS       A36617                    23 bp    DNA     linear   PAT 05-MAR-1997
DEFINITION  Sequence 26 from Patent WO9325707.
ACCESSION   A36617
VERSION     A36617.1  GI:2293921
KEYWORDS    .
SOURCE      Human herpesvirus 3 (HHV-3)
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Tenorio,M.A.
  TITLE     METHODS FOR AMPLIFICATION OF GENOME AND MIXTURES OF INITIATOR
            OLIGONUCLEOTIDES FOR THE DETECTION AND IDENTIFICATION OF RELATED
            GENOMIC SEQUENCES
  JOURNAL   Patent: WO 9325707-A 26 23-DEC-1993;
            INST DE SALUD CARLOS III (ES)
COMMENT     Other publication AU 4327593 940104
            Other publication ES 2048652 940316.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Human herpesvirus 3"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10335"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 6
G32838
LOCUS       G32838                   104 bp    DNA     linear   STS 24-SEP-1999
DEFINITION  A009U15 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G32838
VERSION     G32838.1  GI:5923359
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 104)
  AUTHORS   Adams,M.D.
  TITLE     Human STS sequences
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Email: mdadams@tigr.org
            Primer A: CAGCTTTCTACTTTGACAATGTA
            Primer B: ACAGGAGGTTATAGACAACAGTT
            STS size: 104
            PCR Profile:
                      Denaturation:  96C 5min
                      Anneal:        54C 30sec
                      Extend:        72C 30sec
                      Denature:      95C 30sec
                      FinalExtend:   72C  5min
                      Cycles:        30
```

```
           Protocol:
                    GenomicDNA:   25 ng
                    Primer:       0.43 uM each
                    dNTPs:       230 uM each
                    AmpliTaq:     0.5 units
                    TaqStart Ab:  0.5 units
                    Total Volume: 10 ul

           Buffer:
                    Tris-HCl pH8.8: 100 mM
                    KCl:            500 mM
                    MgCl2:           20 mM
                    Triton X-100:    1%
                    Concentration:  10X

           Prepared with primer pairs derived from THC131873: GenBank
           Accession Numbers-  R85291, T97765, L75280.
FEATURES             Location/Qualifiers
     source          1. .104
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human"
     STS             1. .104
     primer_bind     1. .23
     primer_bind     complement(82. .104)
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 104;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         62 AAGGTTATATAT 73


RESULT 7
CR378386/c
LOCUS       CR378386                113 bp    DNA     linear   STS 24-MAR-2004
DEFINITION  Arabidopsis thaliana transposon insertion STS GT_3.5710, sequence
            tagged site.
ACCESSION   CR378386
VERSION     CR378386.1  GI:45724825
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B., Walsh,S.,
            Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2004) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N174445.
FEATURES             Location/Qualifiers
     source          1. .113
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="DSG6 x AC4"
                     /db_xref="taxon:3702"
                     /clone="AL079349"
                     /ecotype="Landsberg erecta NASC stock code NW20"
     STS             1. .113
                     /standard_name="GT_3.5710"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         15 AAGGTTATATAT 4


RESULT 8
BX294638
LOCUS       BX294638                139 bp    DNA     linear   STS 11-JUN-2003
DEFINITION  Arabidopsis thaliana transposon insertion STS SM_3.3404, sequence
            tagged site.
ACCESSION   BX294638
VERSION     BX294638.1  GI:29125649
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B.,
            Murphy,G., Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N57315.
FEATURES             Location/Qualifiers
     source          1. .139
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia-0 NASC stock code N1092"
                     /db_xref="taxon:3702"
                     /clone="AC002062"
                     /note="Derived from superpool NASC code Unknown"
     STS             1. .139
                     /standard_name="SM_3.3404"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 139;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         95 AAGGTTATATAT 106


RESULT 9
BD043274/c
LOCUS       BD043274                146 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD043274
```

```
VERSION     BD043274.1  GI:22585016
KEYWORDS    JP 2001269182-A/19520.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 19520 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/19520
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .146
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 10
AX907741/c
LOCUS       AX907741                146 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 23604 from Patent EP1033401.
ACCESSION   AX907741
VERSION     AX907741.1  GI:40063821
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE     Expressed sequence tags and encoded human proteins
  JOURNAL   Patent: EP 1033401-A 23604 06-SEP-2000;
            Genset (FR)
FEATURES              Location/Qualifiers
     source           1. .146
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 11
AF487987/c
LOCUS       AF487987                169 bp    DNA     linear   INV 09-OCT-2002
DEFINITION  Triatoma dimidiata clone TDMS16, microsatellite sequence.
ACCESSION   AF487987
VERSION     AF487987.1  GI:19744817
KEYWORDS    .
SOURCE      Triatoma dimidiata (kissing bug)
  ORGANISM  Triatoma dimidiata
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Heteroptera;
            Panheteroptera; Cimicomorpha; Reduviidae; Triatominae; Triatoma.
REFERENCE   1  (bases 1 to 169)
  AUTHORS   Anderson,J.M., Lai,J.E., Dotson,E.M., Cordon-Rosales,C., Ponce,C.,
            Norris,D.E. and Beard,C.B.
  TITLE     Identification and characterization of microsatellite markers in
            the Chagas disease vector Triatoma dimidiata
  JOURNAL   Infect. Genet. Evol. 1 (3), 243-248 (2002)
REFERENCE   2  (bases 1 to 169)
  AUTHORS   Anderson,J.M., Lai,J.E., Dotson,E.M. and Beard,C.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) NCID/DPD/Entomology Branch, Centers for
            Disease Control and Prevention, 4770 Buford Hwy, Mail Stop F-22,
            Atlanta, GA 30341-3724, USA
FEATURES              Location/Qualifiers
     source           1. .169
                      /organism="Triatoma dimidiata"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:72491"
                      /clone="TDMS16"
     repeat_region    1. .169
                      /note="microsatellite"
                      /rpt_type=tandem
ORIGIN

  Query Match           100.0%;  Score 12;  DB 2;  Length 169;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        129 AAGGTTATATAT 118


RESULT 12
BD268181/c
LOCUS       BD268181                184 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Gene to be the target for herbicide and method for targeting the
            gene.
ACCESSION   BD268181
VERSION     BD268181.1  GI:33077949
KEYWORDS    JP 2002534128-A/8.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Levin,J.Z., Budziszewski,G.J., Louis,S.L.P. and Glover,L.W.
  TITLE     Gene to be the target for herbicide and method for targeting the
  JOURNAL   Patent: JP 2002534128-A 8 15-OCT-2002;
            SYNGENTA PARTICIPATIONS AG
COMMENT     OS   Arabidopsis thaliana (thale cress)
            PN   JP 2002534128-A/8
            PD   15-OCT-2002
            PF   13-JAN-2000 JP 2000593762
            PR   15-JAN-1999 US    09/232760,26-JAN-1999 US    09/237479 PR
            03-FEB-1999 US    09/244288,18-FEB-1999 US    09/252336 PR
            30-MAR-1999 US    09/281376
            PI   JOSHUA ZVI LEVIN,GREGORY JOSEPH BUDZISZEWSKI,SHARON LEE POTTER

            PI   LOUIS,
```

```
            PI   LYN WEGRICH GLOVER
            PC   C12N15/09,A01H5/00,C12N1/19,C12N1/21,C12N5/10,C12Q1/02 PC
            ,G01N33/15,G01N33/50,
            PC   G01N33/566,C12N15/00,C12N5/00
            CC   Gene to be the target for herbicide and method for targeting
            CC             the gene
            FH   Key             Location/Qualifiers
            FT   source          1. .184
            FT                   /organism='Arabidopsis thaliana (thale  FT
                        cress)'.
FEATURES              Location/Qualifiers
     source           1. .184
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:3702"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 13
AR210010/c
LOCUS       AR210010                184 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 13 from patent US 6387637.
ACCESSION   AR210010
VERSION     AR210010.1  GI:21512129
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Levin,J.Z., Budziszewski,G.J., Potter,S.L. and Wegrich,L.M.
  TITLE     Herbicide target genes and method
  JOURNAL   Patent: US 6387637-A 13 14-MAY-2002;
FEATURES              Location/Qualifiers
     source           1. .184
                      /organism="unknown"
                      /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 14
AX196591
LOCUS       AX196591                188 bp    DNA     linear   PAT 07-SEP-2001
DEFINITION  Sequence 298 from Patent WO0151627.
ACCESSION   AX196591
VERSION     AX196591.1  GI:15386797
KEYWORDS    .
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1
  AUTHORS   Hauge,B.M., Wang,M.L., Parsons,J.D. and Parnell,L.D.
  TITLE     Nucleic acid molecules and other molecules associated with soybean
            cyst nematode resistance
  JOURNAL   Patent: WO 0151627-A 298 19-JUL-2001;
            MONSANTO COMPANY (US)
FEATURES              Location/Qualifiers
     source           1. .188
                      /organism="Glycine max"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:3847"
                      /note="Seq ID: 318013_region_A3__153114_12"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 15
G23922
LOCUS       G23922                  190 bp    DNA     linear   STS 31-MAY-1996
DEFINITION  human STS WI-16913, sequence tagged site.
ACCESSION   G23922
VERSION     G23922.1  GI:1344248
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TACTGGGCTTGCAAAAACCC
            Primer B: GACTTTTGCTGTGCCCCAGT
            STS size: 150
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T97765).
FEATURES              Location/Qualifiers
     source           1. .190
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
```

```
                     /map="338.9 cR from top of Chr17 linkage group"
     STS             1. .150
     primer_bind     1. .20
     primer_bind     complement(131. .150)
ORIGIN

  Query Match            100.0%;  Score 12;  DB 10;  Length 190;
  Best Local Similarity  100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         85 AAGGTTATATAT 96


RESULT 16
AB153981/c
LOCUS       AB153981                200 bp    DNA     linear   STS 02-APR-2005
DEFINITION  Homo sapiens DNA, STS on chromosome 11, D11S0972i, sequence tagged
            site.
ACCESSION   AB153981
VERSION     AB153981.1  GI:62174799
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tamiya,G., Shinya,M., Ikuta,T., Makino,S., Matsumoto,T., Mano,S.,
            Furugaki,K., Ando,S., Nozaki,Y., Yukawa,W., Nakashige,R.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Nakami,Y., Takayama,S.,
            Endo,T., Saruwatari,T., Yagura,M., Yoshikawa,Y., Fujimoto,K.,
            Oka,A., Chiku,S., Linsen,S.E., Giphart,M.J., Bahram,S.,
            Kulski,Y.J., Fukazawa,T., Hashimoto,H., Hoshina,Y., Suzuki,Y.,
            Hotta,T., Mochida,J., Minezaki,T., Komai,K., Shiozawa,S.,
            Taniguchi,A., Yamanaka,H., Kamatani,N., Imanishi,T., Gojobori,T.
            and Inoko,H.
  TITLE     A Whole Genome Association Study of Rheumatoid Arthritis using
            27,039 Microsatellite Makers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200)
  AUTHORS   Tamiya,G., Makino,S., Fujimoto,K., Oka,A., Hayashi,H., Denda,A.,
            Linsen,S.E., Ikuta,T., Shinya,M., Endo,T., Tomizawa,M., Tokubo,E.,
            Sato,R., Takaki,A., Nagatsuka,Y., Watanabe,H., Adachi,S.,
            Makino,Y., Nakano,S., Yamamoto,A., Yoshida,K., Okamoto,K.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Takayama,S., Nakami,Y.,
            Saruwatari,T., Brand,A., van Hilten,J.A., van de Watering,L.M.,
            Giphart,M.J., Bahram,S., Kulski,Y.J. and Inoko,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Hidetoshi Inoko, Tokai University School of
            Medicine, Department of Genetic Information; Bohseidai, Isehara,
            Kanagawa, 259-1193, Japan (E-mail:hinoko@is.icc.u-tokai.ac.jp,
            Tel:81-463-93-1121, Fax:81-463-94-8884)
COMMENT     Polymorphisms were confirmed by comparing with the pooled DNA
            typing data of 88 Dutch population or 100 Australian. When there
            was no significant difference between the Japanese and these
            Caucasians, individual typing was performed to confirm
            polymorphisms.
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q14.3"
                     /note="pooled DNA of 100 Japanese unrelated individuals
                     sequence tagged site D11S0972i"
     primer_bind     1. .23
                     /note="5' primer: GGATAAAGTAATAGTGCTTGCAA"
                     /PCR_conditions="denaturation 96degC 5 min, 57degC 1 min,
                     72degC 1 min; 40 cycles 96degC 45 sec, 57degC 45 sec,
                     72degC 1 min"
     repeat_region   67. .177
                     /note="sequence tagged site D11S0972i"
                     /rpt_type=tandem
                     /rpt_unit="ac"
                     /rpt_unit="at"
                     /rpt_unit="ta"
     primer_bind     complement(177. .200)
                     /note="3' primer: CATAGCTACTCTCACATGAAGGTT"
ORIGIN

  Query Match            100.0%;  Score 12;  DB 10;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        182 AAGGTTATATAT 171


RESULT 17
ATH531582
LOCUS       ATH531582               201 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            237F12.
ACCESSION   AJ531582
VERSION     AJ531582.1  GI:26799842
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="237F12"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .201
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match            100.0%;  Score 12;  DB 15;  Length 201;
  Best Local Similarity  100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        116 AAGGTTATATAT 127


RESULT 18
ATH531598
LOCUS       ATH531598                201 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            238A11.
ACCESSION   AJ531598
VERSION     AJ531598.1  GI:26799858
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="238A11"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .201
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        116 AAGGTTATATAT 127


RESULT 19
G31833
LOCUS       G31833                   225 bp    DNA     linear   STS 28-SEP-1998
DEFINITION  SpAGH1 Norway spruce endosperms Picea abies STS genomic, sequence
            tagged site.
ACCESSION   G31833
VERSION     G31833.1  GI:1927164
KEYWORDS    STS.
SOURCE      Picea abies (Norway spruce)
  ORGANISM  Picea abies
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Pfeiffer,A., Olivieri,A.M. and Morgante,M.
  TITLE     Identification and characterization of microsatellites in Norway
            spruce (Picea abies K.)
  JOURNAL   Genome 40 (4), 411-419 (1997)
   PUBMED   9276931
COMMENT
            Contact: Michele Morgante
            Dipartimento di Produzione Vegetale e Tecnologie Agrarie
            Universita' degli Studi di Udine
            Via delle Scienze 208, I-33100 Udine, ITALY
            Tel: 39432558606
            Fax: 39432558603
            Email: morgante@dpvta.uniud.it
            Primer A: Unknown
            Primer B: Unknown.
FEATURES             Location/Qualifiers
     source          1. .225
                     /organism="Picea abies"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3329"
                     /tissue_type="Endosperm"
                     /clone_lib="Norway spruce endosperms"
                     /note="Vector: lambda ZAP II; V-type: Phage; Size selected
                     (200-600 bp) genomic fragments obtained through
                     restriction enzyme digestion cloned into lambda ZapII"
     STS             <1. .>225
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        117 AAGGTTATATAT 128


RESULT 20
AY020751/c
LOCUS       AY020751                 226 bp    DNA     linear   PLN 07-FEB-2001
DEFINITION  Oryza sativa microsatellite MRG3076 containing (TA)X13, genomic
            sequence.
ACCESSION   AY020751
VERSION     AY020751.1  GI:12703793
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 226)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Simple sequence repeats from Monsanto rice genomic sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
            Blvd., Creve Coeur, MO 63167, USA
COMMENT     Derived from rice genomic sequences generated from the Monsanto
            Rice Genome Sequencing project. Please see
            http://www.rice-research.org for more information.  The sequence
            data were produced primarily in the laboratories of Dr. Leroy Hood
            at the University of Washington in Seattle.
FEATURES             Location/Qualifiers
     source          1. .226
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
```

```
     repeat_region   1. .226
                     /note="microsatellite MRG3076"
                     /rpt_type=tandem
                     /rpt_unit="ta"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        131 AAGGTTATATAT 120


RESULT 21
MMU409475
LOCUS       MMU409475               227 bp    mRNA    linear   ROD 27-JUN-2001
DEFINITION  Mus musculus RNA binding site for Dazl protein, clone cg10.
ACCESSION   AJ409475
VERSION     AJ409475.1  GI:13016754
KEYWORDS    Dazl gene; RNA binding protein.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   Venables,J.P., Ruggiu,M. and Cooke,H.J.
  TITLE     The RNA-binding specificity of the mouse Dazl protein
  JOURNAL   Nucleic Acids Res. 29 (12), 2479-2483 (2001)
   PUBMED   11410654
REFERENCE   2  (bases 1 to 227)
  AUTHORS   Venables,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2001) Venables J.P., Chromosome Biology, MRC
            Human Gentics Unit, Western Genral Hospital, Crewe Rd, Edinburgh
            EH4 2XU, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .227
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="cg10"
     protein_bind    1. .227
                     /note="U-rich, Dazl binding site"
                     /bound_moiety="RNA binding protein, Dazl"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 9;  Length 227;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        207 AAGGTTATATAT 218


RESULT 22
AB143702/c
LOCUS       AB143702                232 bp    DNA     linear   STS 02-APR-2005
DEFINITION  Homo sapiens DNA, STS on chromosome 20, D20S0259i, sequence tagged
            site.
ACCESSION   AB143702
VERSION     AB143702.1  GI:62164520
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tamiya,G., Shinya,M., Ikuta,T., Makino,S., Matsumoto,T., Mano,S.,
            Furugaki,K., Ando,S., Nozaki,Y., Yukawa,W., Nakashige,R.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Nakami,Y., Takayama,S.,
            Endo,T., Saruwatari,T., Yagura,M., Yoshikawa,Y., Fujimoto,K.,
            Oka,A., Chiku,S., Linsen,S.E., Giphart,M.J., Bahram,S.,
            Kulski,Y.J., Fukazawa,T., Hashimoto,H., Hoshina,Y., Suzuki,Y.,
            Hotta,T., Mochida,J., Minezaki,T., Komai,K., Shiozawa,S.,
            Taniguchi,A., Yamanaka,H., Kamatani,N., Imanishi,T., Gojobori,T.
            and Inoko,H.
  TITLE     A Whole Genome Association Study of Rheumatoid Arthritis using
            27,039 Microsatellite Makers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232)
  AUTHORS   Tamiya,G., Makino,S., Fujimoto,K., Oka,A., Hayashi,H., Denda,A.,
            Linsen,S.E., Ikuta,T., Shinya,M., Endo,T., Tomizawa,M., Tokubo,E.,
            Sato,R., Takaki,A., Nagatsuka,Y., Watanabe,H., Adachi,S.,
            Makino,Y., Nakano,S., Yamamoto,A., Yoshida,K., Okamoto,K.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Takayama,S., Nakami,Y.,
            Saruwatari,T., Brand,A., van Hilten,J.A., van de Watering,L.M.,
            Giphart,M.J., Bahram,S., Kulski,Y.J. and Inoko,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Hidetoshi Inoko, Tokai University School of
            Medicine, Department of Genetic Information; Bohseidai, Isehara,
            Kanagawa, 259-1193, Japan (E-mail:hinoko@is.icc.u-tokai.ac.jp,
            Tel:81-463-93-1121, Fax:81-463-94-8884)
COMMENT     Polymorphisms were confirmed by comparing with the pooled DNA
            typing data of 88 Dutch population or 100 Australian. When there
            was no significant difference between the Japanese and these
            Caucasians, individual typing was performed to confirm
            polymorphisms.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20p12.3"
                     /note="pooled DNA of 100 Japanese unrelated individuals
                     sequence tagged site D20S0259i"
     primer_bind     1. .24
                     /note="5' primer: CAAACATTGAACCATTACTTCTAG"
                     /PCR_conditions="denaturation 96degC 5 min, 57degC 1 min,
                     72degC 1 min; 40 cycles 96degC 45 sec, 57degC 45 sec,
                     72degC 1 min"
     repeat_region   128. .202
                     /note="sequence tagged site D20S0259i"
                     /rpt_type=tandem
                     /rpt_unit="ta"
                     /rpt_unit="tatatg"
     primer_bind     complement(210. .232)
                     /note="3' primer: GGTCATGAGTTCAAGGTTAATAC"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 85


RESULT 23
AY020956
LOCUS       AY020956                232 bp    DNA     linear   PLN 07-FEB-2001
DEFINITION  Oryza sativa microsatellite MRG3281 containing (TA)X16, genomic
            sequence.
ACCESSION   AY020956
VERSION     AY020956.1  GI:12703998
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Simple sequence repeats from Monsanto rice genomic sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
            Blvd., Creve Coeur, MO 63167, USA
COMMENT     Derived from rice genomic sequences generated from the Monsanto
            Rice Genome Sequencing project. Please see
            http://www.rice-research.org for more information.  The sequence
            data were produced primarily in the laboratories of Dr. Leroy Hood
            at the University of Washington in Seattle.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
     repeat_region   1. .232
                     /note="microsatellite MRG3281"
                     /rpt_type=tandem
                     /rpt_unit="ta"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 107


RESULT 24
AY114172
LOCUS       AY114172                234 bp    DNA     linear   VRL 26-FEB-2003
DEFINITION  Human herpesvirus 3 DNA polymerase gene, partial cds.
ACCESSION   AY114172
VERSION     AY114172.2  GI:28566755
KEYWORDS    .
SOURCE      Human herpesvirus 3 (HHV-3)
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Detection and typing of human herpesviruses 1-5 by consensus primer
            PCR and RFLP
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2002) Pediatric, Chulalongkorn University, Rama
            IV, Patumwan, Bangkok 10330, Thailand
REFERENCE   3  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2003) Pediatric, Chulalongkorn University, Rama
            IV, Patumwan, Bangkok 10330, Thailand
  REMARK    Sequence update by submitter
COMMENT     On Feb 26, 2003 this sequence version replaced gi:21724220.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Human herpesvirus 3"
                     /proviral
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10335"
                     /note="VZV"
     CDS             <1. .>234
                     /codon_start=1
                     /product="DNA polymerase"
                     /protein_id="AAM75496.2"
                     /db_xref="GI:28566756"
                     /translation="CNSVYGFTGVAQGFLPCLYVAATVTTIGRQMLLSTRDYIHNNWA
                     APERFITAFPDIESSVLSQKAYEVKVIYGDTDSV"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 13;  Length 234;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        205 AAGGTTATATAT 216


RESULT 25
G37836/c
LOCUS       G37836                  236 bp    DNA     linear   STS 24-MAR-1999
DEFINITION  wh_hsp1 Plasmodium falciparum haploid Plasmodium falciparum STS
            genomic, sequence tagged site.
ACCESSION   G37836
VERSION     G37836.1  GI:3015442
KEYWORDS    STS.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Su,X.-z., Ferdig,M.T. and Wellems,T.E.
  TITLE     Integrated genetic and physical maps of Plasmodium falciparum
  JOURNAL   Unpublished (1998)
COMMENT     Synonyms: wh_hsp1
            Contact: Thomas E. Wellems
            LPD
            NIAID
            Bldg. 4, Room 126, NIH Campus, Bethesda, MD 20892-0425
            Tel: (301) 496-4021
            Fax: (301) 402-0079
            Email: tew@helix.nih.gov
            Primer A: GCATATTATTATGAATTACC
            Primer B: ATCAAAAGGTTATATATGTA
            STS size: 236
            PCR Profile:
            Initial heat: 2min at 94oC
            Cycles:       30 cycles, each having steps 20sec at 94oC;
                          10sec at 45oC; 10sec at 40oC; 30sec at 60oC
            Machine:      Perkin Elmer GeneAmp 9600
            Protocol:
                  Template:        1-10 ng
                  Primer:          5 pM each
                  dNTPs:           0.3 ul of 10 mM stock solution
                  Taq Polymerase:  0.5 unit
                  Total Volume:    15 ul

            Buffer:
                  10 mM Tris HCl, pH 8.3
                  1.5 mM MgCl2
                  50 mM KCl.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3xDd2"
                     /db_xref="taxon:5833"
                     /map="9"
                     /clone_lib="Plasmodium falciparum haploid"
                     /dev_stage="erythrocytic haploid stage"
     STS             1. .236
     primer_bind     1. .20
     primer_bind     complement(217. .236)
ORIGIN
```

```
  Query Match             100.0%;  Score 12;  DB 10;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        231 AAGGTTATATAT 220


RESULT 26
G44501/c
LOCUS       G44501                   236 bp    DNA     linear   STS 24-MAR-1999
DEFINITION  C9M106 Plasmodium falciparum haploid Plasmodium falciparum STS
            genomic, sequence tagged site.
ACCESSION   G44501
VERSION     G44501.2  GI:4511820
KEYWORDS    STS.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Su,X.-z., Ferdig,M.T. and Wellems,T.E.
  TITLE     Integrated genetic and physical maps of Plasmodium falciparum
  JOURNAL   Unpublished (1998)
COMMENT     On Mar 25, 1999 this sequence version replaced gi:4218688.

            Contact: Thomas E. Wellems
            LPD
            NIAID
            Bldg. 4, Room 126, NIH Campus, Bethesda, MD 20892-0425
            Tel: (301) 496-4021
            Fax: (301) 402-0079
            Email: tew@helix.nih.gov
            Primer A: CATATTTATTTATTCATGTCC
            Primer B: ATCAAAAGGTTATATATGTA
            STS size: 236
            PCR Profile:
            Initial heat: 2min at 94oC
            Cycles:       30 cycles, each having steps 20sec at 94oC;
                          10sec at 45oC; 10sec at 40oC; 30sec at 60oC
            Machine:      Perkin Elmer GeneAmp 9600
            Protocol:
                  Template:        1-10 ng
                  Primer:          5 pM each
                  dNTPs:           0.3 ul of 10 mM stock solution
                  Taq Polymerase:  0.5 unit
                  Total Volume:    15 ul

            Buffer:
                  10 mM Tris HCl, pH 8.3
                  1.5 mM MgCl2
                  50 mM KCl.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3xDd2"
                     /db_xref="taxon:5833"
                     /map="9"
                     /clone_lib="Plasmodium falciparum haploid"
                     /dev_stage="erythrocytic haploid stage"
     STS             1. .236
     primer_bind     1. .21
     primer_bind     complement(217. .236)
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        231 AAGGTTATATAT 220


RESULT 27
AU026951/c
LOCUS       AU026951                 238 bp    DNA     linear   STS 02-MAR-1999
DEFINITION  Rattus norvegicus, OTSUKA clone, OT62.26/819g05, microsatellite
            sequence, sequence tagged site.
ACCESSION   AU026951
VERSION     AU026951.1  GI:4516874
KEYWORDS    STS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A.,
            Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M.,
            Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,T., Nakamura,Y. and
            Tanigami,A.
  TITLE     The large-scale mapping of rat microsatellite markers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238)
  AUTHORS   Watanabe,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) Takeshi K Watanabe, Otsuka Pharmaceutical
            Co., Ltd., Otsuka GEN Research Institute; 463-10, Kagasuno,
            Kawauchi-cho, Tokushima, Tokushima 771-0192, Japan
            (E-mail:watanabe@otsuka.genome.ad.jp, Tel:81-886-65-2888,
            Fax:81-886-37-1035)
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="OT62.26/819g05"
                     /note="OT62.26/819g05F=5'-ATGCGACAGAAGCCAGAAAC-3',
                     OT62.26/819g05R=5'-CAGTTAGCCTTGAGCTTAAAGCA-3'"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        209 AAGGTTATATAT 198


RESULT 28
BD028743/c
LOCUS       BD028743                 242 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD028743
VERSION     BD028743.1  GI:22570485
KEYWORDS    JP 2001269182-A/4989.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 4989 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/4989
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
```

```
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .242
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 29
AX893210/c
LOCUS       AX893210                242 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 9073 from Patent EP1033401.
ACCESSION   AX893210
VERSION     AX893210.1  GI:40048094
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE     Expressed sequence tags and encoded human proteins
  JOURNAL   Patent: EP 1033401-A 9073 06-SEP-2000;
            Genset (FR)
FEATURES              Location/Qualifiers
     source           1. .242
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 30
CQ457358
LOCUS       CQ457358                255 bp    DNA     linear   PAT 30-JAN-2004
DEFINITION  Sequence 136 from Patent WO0192581.
ACCESSION   CQ457358
VERSION     CQ457358.1  GI:41422977
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Algate,P.A., Harlocker,S.L. and Jones,R.
  TITLE     Compositions and methods for the therapy and diagnosis of
            ovariancancer
  JOURNAL   Patent: WO 0192581-A 136 06-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES              Location/Qualifiers
     source           1. .255
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37


Search completed: December 20, 2005, 10:43:04
Job time : 72.5382 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:22:54 ; Search time 9.22892 Seconds
(without alignments)
8665.841 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	13	5	ABF93204	Abf93204 Oligonucl
c	2	12	100.0	13	5	ABF93205	Abf93205 Oligonucl
	3	12	100.0	23	2	AAQ53332	Aaq53332 Positive
	4	12	100.0	23	2	AAT37524	Aat37524 Varicella
	5	12	100.0	100	12	ACH88864	Ach88864 Human gen
	6	12	100.0	121	11	ADZ42781	Adz42781 Human gen
c	7	12	100.0	146	3	AAC19529	Aac19529 Human sec
	8	12	100.0	146	12	ACH81646	Ach81646 Human gen
c	9	12	100.0	184	3	AAA50085	Aaa50085 Arabidops
	10	12	100.0	188	5	AAI61667	Aai61667 Soybean 3
c	11	12	100.0	242	3	AAC04998	Aac04998 Human sec
	12	12	100.0	255	6	ABL77158	Abl77158 Human ova
	13	12	100.0	267	6	ABN77348	Abn77348 Human ORF
c	14	12	100.0	277	5	ABV07118	Abv07118 Human pro
c	15	12	100.0	289	6	ABL77086	Abl77086 Human ova
c	16	12	100.0	294	3	AAC18502	Aac18502 Human sec
c	17	12	100.0	313	6	ABL80474	Abl80474 Human ova
	18	12	100.0	316	6	ABK46190	Abk46190 cDNA enco
	19	12	100.0	317	4	AAH69998	Aah69998 Human cer
	20	12	100.0	317	4	AAH73195	Aah73195 Human cer
	21	12	100.0	320	4	AAH70509	Aah70509 Human cer
	22	12	100.0	341	4	AAH69769	Aah69769 Human cer
	23	12	100.0	362	3	AAC24388	Aac24388 Human sec
	24	12	100.0	366	4	AAH71423	Aah71423 Human cer
	25	12	100.0	367	4	AAH35822	Aah35822 Human col
	26	12	100.0	374	4	AAS37872	Aas37872 Novel hum
c	27	12	100.0	375	13	ADV83971	Adv83971 Streptoco
	28	12	100.0	378	4	AAK81895	Aak81895 Human imm
	29	12	100.0	378	4	AAK81894	Aak81894 Human imm
c	30	12	100.0	378	4	AAK62195	Aak62195 Human imm
	31	12	100.0	385	10	ABT22768	Abt22768 Breast ca
c	32	12	100.0	402	4	AAK87775	Aak87775 Human dig
c	33	12	100.0	402	5	AAS31692	Aas31692 Human liv
c	34	12	100.0	402	6	ABN90047	Abn90047 Human liv
c	35	12	100.0	402	11	ADJ14814	Adj14814 Human liv
c	36	12	100.0	405	4	AAK60591	Aak60591 Human imm
	37	12	100.0	405	9	ACH18939	Ach18939 Human adu
	38	12	100.0	410	8	ABX44501	Abx44501 Bovine ES
c	39	12	100.0	411	10	ADI21471	Adi21471 Novel hum
	40	12	100.0	412	6	ABK64663	Abk64663 Human ben
	41	12	100.0	420	14	AEB90163	Aeb90163 Isolated
c	42	12	100.0	421	8	ABZ52835	Abz52835 Aspergill
c	43	12	100.0	422	4	AAI87578	Aai87578 Human pol
c	44	12	100.0	423	13	ACF85182	Acf85182 Human SIR
c	45	12	100.0	432	5	ABV37059	Abv37059 Human pro

ALIGNMENTS

RESULT 1
ABF93204
ID ABF93204 standard; DNA; 13 BP.
XX
AC ABF93204;
XX
DT 22-FEB-2002 (first entry)
XX
DE Oligonucleotide SEQ ID NO 193201 for detecting SNP TSC0047525.
XX
KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS Homo sapiens.
XX
PN WO200177384-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-IB000713.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-657177/75.
XX
PT Set of oligonucleotides, useful for diagnosis and cell typing, is
PT designed to detect single-nucleotide polymorphisms and cytosine
PT methylation status.
XX
PS Claim 1; SEQ ID NO 193201; 29pp + Sequence Listing; German.
XX
CC This invention describes novel oligonucleotide primers or peptide nucleic
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC and cytosine methylation status in chemically pretreated genomic DNA. The
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC range of diseases including immune system, gastrointestinal, respiratory,
CC central nervous system, cardiovascular and metabolic disorders. The

```
CC   oligomers are also used for detecting cell type differentiation. ABC00010
CC   -ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and ABI00010-ABI82073
CC   represent the oligomers described in the invention. NOTE: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13 BP; 5 A; 0 C; 2 G; 5 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 12;  DB 5;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
ABF93205/c
ID   ABF93205 standard; DNA; 13 BP.
XX
AC   ABF93205;
XX
DT   22-FEB-2002  (first entry)
XX
DE   Oligonucleotide SEQ ID NO 193202 for detecting SNP TSC0047525.
XX
KW   SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW   peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW   central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS   Homo sapiens.
XX
PN   WO200177384-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-IB000713.
XX
PR   07-APR-2000; 2000DE-01019173.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-657177/75.
XX
PT   Set of oligonucleotides, useful for diagnosis and cell typing, is
PT   designed to detect single-nucleotide polymorphisms and cytosine
PT   methylation status.
XX
PS   Claim 1; SEQ ID NO 193202; 29pp + Sequence Listing; German.
XX
CC   This invention describes novel oligonucleotide primers or peptide nucleic
CC   acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC   and cytosine methylation status in chemically pretreated genomic DNA. The
CC   oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC   range of diseases including immune system, gastrointestinal, respiratory,
CC   central nervous system, cardiovascular and metabolic disorders. The
CC   oligomers are also used for detecting cell type differentiation. ABC00010
CC   -ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and ABI00010-ABI82073
CC   represent the oligomers described in the invention. NOTE: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13 BP; 5 A; 2 C; 0 G; 5 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 12;  DB 5;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         13 AAGGTTATATAT 2


RESULT 3
AAQ53332
ID   AAQ53332 standard; DNA; 23 BP.
XX
AC   AAQ53332;
XX
DT   25-MAR-2003  (revised)
DT   30-JUN-1994  (first entry)
XX
DE   Positive primer for detection of VZV.
XX
KW   Varicella Zoster Virus; Herpes viridae; amplification reaction;
KW   DNA polymerase; conserved sequence; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .6
FT                   /*tag=  a
FT                   /note= "correponds to non-conserved region of herpes
FT                   viridae DNA polymerase"
FT   misc_feature    7. .23
FT                   /*tag=  b
FT                   /note= "corresponds to conserved region of herpes viridae
FT                   DNA polymerase"
XX
PN   WO9325707-A2.
XX
PD   23-DEC-1993.
XX
PF   04-JUN-1993;   93WO-ES000048.
XX
PR   05-JUN-1992;   92ES-00001174.
XX
PA   (SALU-) INST SALUD CARLOS III.
XX
PI   Tenorio Matanzo A;
XX
DR   WPI; 1994-007564/01.
XX
PT   Amplification of genome(s) and initiator oligo-nucleotide(s) mixts. -
PT   using single reaction mixt. to detect and identify infections by related
PT   viruses.
XX
PS   Claim 14(i); Page 26; 41pp; Spanish.
XX
CC   The known amino acid sequences of DNA polymerases from 6 different
CC   members of the herpes viridae family were aligned so that conserved
CC   regions could be identified. A set of primers (AAQ53331-5) was designed
CC   based on the conserved amino acid motif IYGDTD and having 6 non-conserved
CC   nucleotides at the 5'-end. Sequence AAQ53332 is specific for VZV.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 23 BP; 9 A; 1 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
AAT37524
ID   AAT37524 standard; DNA; 23 BP.
XX
```

AC AAT37524;
XX
DT 23-APR-1997 (first entry)
XX
DE Varicella-zoster virus genomic amplification primer.
XX
KW Polymerase chain reaction; PCR primer; Herpesviridae; Picornaviridae;
KW human herpesvirus; enterovirus; diagnosis; detection; ss.
XX
OS Varicella-zoster virus.
XX
PN WO9625909-A2.
XX
PD 29-AUG-1996.
XX
PF 16-FEB-1996; 96WO-ES000031.
XX
PR 17-FEB-1995; 95ES-00000320.
XX
PA (SALU-) INST SALUD CARLOS III.
XX
PI Casas I, Tenorio A, Echevarria JM, Klapper PE, Cleator GM;
XX
DR WPI; 1996-402109/40.
XX
PT Differential diagnosis of infectious agents by genomic amplification -
PT using primer mixts., each specific for a family of infectious agents,
PT esp. for distinguishing human herpes virus and entero-virus infection.
XX
PS Claim 13; Page 29; 55pp; Spanish.
XX
CC A mixture of amplification primers was designed which could amplify both
CC enterovirus and herpesvirus sequences from the same sample. The mixture
CC contained at least one of the primers in AAT37520-T37522 and at least one
CC of the primers in AAT37523-T37532. A second PCR can be performed on the
CC products of the first reaction in order to distinguish specific viruses.
CC Alternatively, the PCR products can be detected by hybridisation to virus
CC -specific probes
XX
SQ Sequence 23 BP; 9 A; 1 C; 7 G; 6 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 5e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 1 AAGGTTATATAT 12

RESULT 5
ACH88864
ID ACH88864 standard; DNA; 100 BP.
XX
AC ACH88864;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon probe #22059.
XX
KW Human; probe; ss; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX
DR WPI; 2004-119264/12.
XX
PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
PS Claim 1; SEQ ID NO 22059; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above- mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 100 BP; 31 A; 21 C; 27 G; 21 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 44 AAGGTTATATAT 55

RESULT 6
ADZ42781
ID ADZ42781 standard; DNA; 121 BP.
XX
AC ADZ42781;
XX
DT 16-JUN-2005 (first entry)
XX
DE Human gene encoding a receptor protein with an A61G SNP Seq 363.
XX
KW SNP detection; DNA amplification; DNA microarray;

```
KW   single nucleotide polymorphism; SNP; selectable marker; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       61
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   JP2003274966-A.
XX
PD   30-SEP-2003.
XX
PF   25-MAR-2002; 2002JP-00084465.
XX
PR   25-MAR-2002; 2002JP-00084465.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2003-869435/81.
XX
PT   New polynucleotide isolated from human gene useful for detection of
PT   single nucleotide polymorphisms.
XX
PS   Claim 1; SEQ ID NO 363; 687pp; Japanese.
XX
CC   This invention relates to novel polynucleotides isolated and purified
CC   from 99 human genes that comprises any one of 863 fully defined sequences
CC   of 121 nucleotides, as given in specification. Specifically, it refers to
CC   these polynucleotides as each containing a single base substitution
CC   (otherwise known as a single nucleotide polymorphism) that are useful as
CC   genetic markers and for diagnosing certain diseases. The present
CC   invention provides oligonucleotides containing these SNPs that encode
CC   receptor proteins, as well as PCR primers, labeling probes and a
CC   microarray. Accordingly, this array is useful for diagnosing disease and
CC   determining side effects of a medical agent. This polynucleotide sequence
CC   is a human gene containing a SNP, as given in an exemplification of the
CC   invention.
XX
SQ   Sequence 121 BP; 32 A; 17 C; 24 G; 47 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 12;  DB 11;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         33 AAGGTTATATAT 44


RESULT 7
AAC19529/c
ID   AAC19529 standard; cDNA; 146 BP.
XX
AC   AAC19529;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 23604.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 23604; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 146 BP; 38 A; 44 C; 30 G; 34 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 3;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 8
ACH81646
ID   ACH81646 standard; DNA; 146 BP.
XX
AC   ACH81646;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #14841.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
```

```
XX
PS   Claim 1; SEQ ID NO 14841; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 146 BP; 41 A; 33 C; 37 G; 35 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 12;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         97 AAGGTTATATAT 108


RESULT 9
AAA50085/c
ID   AAA50085 standard; cDNA; 184 BP.
XX
AC   AAA50085;
XX
DT   25-OCT-2000  (first entry)
XX
DE   Arabidopsis herbicide target gene 245 3' UTR cDNA.
XX
KW   Herbicide; 245 gene; inhibitor; screening; transgenic plant;
KW   herbicide resistance; crop protection; peptide release factor 2; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200042205-A2.
XX
PD   20-JUL-2000.
XX
PF   13-JAN-2000; 2000WO-EP000246.
XX
PR   15-JAN-1999;   99US-00232760.
PR   26-JAN-1999;   99US-00237479.
PR   03-FEB-1999;   99US-00244288.
PR   18-FEB-1999;   99US-00252336.
PR   30-MAR-1999;   99US-00281376.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Levin JZ,  Budziszewski GJ,  Potter SL,  Wegrich LM;
XX
DR   WPI; 2000-482835/42.
XX
PT   New DNA, comprising a sequence that codes for a protein essential for
PT   seedling growth, for identifying potential herbicides and developing
PT   herbicide tolerant plants, tissues, seeds and cells.
XX
PS   Claim 46; Page 87; 100pp; English.
XX
CC   This is the 3' untranslated region (UTR) of cDNA corresponding to the
CC   Arabidopsis thaliana 245 gene. The coding region (see AAA50078) of the
CC   cDNA codes for a protein (see AAY95654) that is essential for normal
CC   seedling growth and development, and which shows sequence similarity to
CC   peptide release factor 2. The 245 cDNA was isolated from a cDNA library
CC   using Arabidopsis peptide release factor DNA as probe. The 245 protein is
CC   a useful herbicide target, and the invention provides screening assays
CC   useful for identifying inhibitors that are potential herbicides. The
CC   invention is also applied to using the 245 gene for the development of
CC   herbicide tolerant plants, plant tissues, plant seeds, and plant cells
XX
SQ   Sequence 184 BP; 84 A; 23 C; 25 G; 50 T; 0 U; 2 Other;

  Query Match             100.0%;  Score 12;  DB 3;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 10
AAI61667
ID   AAI61667 standard; DNA; 188 BP.
XX
AC   AAI61667;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 318013 region A3 DNA, SEQ ID NO: 298.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US000552.
XX
PR   07-JAN-2000; 2000US-0174880P.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs.
XX
PS   Claim 25; Page 1105; 1353pp; English.
```

```
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule provided
CC   in the specification
XX
SQ   Sequence 188 BP; 79 A; 20 C; 16 G; 73 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 5;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 11
AAC04998/c
ID   AAC04998 standard; cDNA; 242 BP.
XX
AC   AAC04998;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 9073.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 9073; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 242 BP; 55 A; 50 C; 25 G; 100 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 12;  DB 3;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 12
ABL77158
ID   ABL77158 standard; cDNA; 255 BP.
XX
AC   ABL77158;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:136.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 136; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 255 BP; 88 A; 53 C; 36 G; 78 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37
```

RESULT 13
ABN77348
ID ABN77348 standard; cDNA; 267 BP.
XX
AC ABN77348;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF2295 cDNA, SEQ ID NO:4589.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shimkets RA;
XX
DR WPI; 2002-106200/14.
DR P-PSDB; ABP33322.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX
PS Claim 1; Page 1401; 2508pp; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 267 BP; 67 A; 52 C; 58 G; 89 T; 0 U; 1 Other;

  Query Match   100.0%; Score 12; DB 6; Length 267;
  Best Local Similarity   100.0%; Pred. No. 4.5e+03;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         23 AAGGTTATATAT 34
```

RESULT 14
ABV07118/c
ID ABV07118 standard; cDNA; 277 BP.
XX
AC ABV07118;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 7109.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 1155; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

```
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 277 BP; 75 A; 51 C; 61 G; 88 T; 0 U; 2 Other;

  Query Match            100.0%;  Score 12;  DB 5;  Length 277;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         76 AAGGTTATATAT 65


RESULT 15
ABL77086/c
ID   ABL77086 standard; cDNA; 289 BP.
XX
AC   ABL77086;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:64.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 64; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 289 BP; 99 A; 36 C; 56 G; 97 T; 0 U; 1 Other;

  Query Match            100.0%;  Score 12;  DB 6;  Length 289;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        250 AAGGTTATATAT 239


RESULT 16
AAC18502/c
ID   AAC18502 standard; cDNA; 294 BP.
XX
AC   AAC18502;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 22577.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 22577; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 294 BP; 98 A; 49 C; 37 G; 107 T; 0 U; 3 Other;

  Query Match            100.0%;  Score 12;  DB 3;  Length 294;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        240 AAGGTTATATAT 229
```

```
RESULT 17
ABL80474/c
ID   ABL80474 standard; cDNA; 313 BP.
XX
AC   ABL80474;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:3452.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 3452; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 313 BP; 75 A; 54 C; 72 G; 106 T; 0 U; 6 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 313;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        303 AAGGTTATATAT 292


RESULT 18
ABK46190
ID   ABK46190 standard; cDNA; 316 BP.
XX
AC   ABK46190;
XX
DT   05-JUN-2002  (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 1741.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US024218.
XX
PR   03-AUG-2000; 2000US-0223283P.
PR   28-MAR-2001; 2001US-0279763P.
PR   29-JUN-2001; 2001US-0302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE,  Meagher MJ,  Xu J,  Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers for
PT   the progression of cancer.
XX
PS   Claim 1; SEQ ID NO 1741; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer. ABK44450-
CC   ABK46237 represent coding sequences of human colon tumour proteins of the
CC   invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC   data for this patent did not form part of the printed specification but
CC   was supplied by the European Patent Office
XX
SQ   Sequence 316 BP; 111 A; 56 C; 69 G; 76 T; 0 U; 4 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        161 AAGGTTATATAT 172


RESULT 19
AAH69998
ID   AAH69998 standard; cDNA; 317 BP.
XX
AC   AAH69998;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1272.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
```

```
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 303; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 317 BP; 79 A; 64 C; 58 G; 89 T; 0 U; 27 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         50 AAGGTTATATAT 61


RESULT 20
AAH73195
ID   AAH73195 standard; cDNA; 317 BP.
XX
AC   AAH73195;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 4469.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 977; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 317 BP; 108 A; 51 C; 59 G; 99 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        265 AAGGTTATATAT 276


RESULT 21
AAH70509
ID   AAH70509 standard; cDNA; 320 BP.
XX
AC   AAH70509;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1783.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 386; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
```

```
SQ   Sequence 320 BP; 108 A; 51 C; 61 G; 100 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 4;  Length 320;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        268 AAGGTTATATAT 279


RESULT 22
AAH69769
ID   AAH69769 standard; cDNA; 341 BP.
XX
AC   AAH69769;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1043.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 272; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 341 BP; 110 A; 58 C; 69 G; 104 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 4;  Length 341;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        290 AAGGTTATATAT 301


RESULT 23
AAC24388
ID   AAC24388 standard; cDNA; 362 BP.
XX
AC   AAC24388;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 28463.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 28463; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 362 BP; 105 A; 74 C; 110 G; 73 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 3;  Length 362;
  Best Local Similarity  100.0%;  Pred. No. 4.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        295 AAGGTTATATAT 306


RESULT 24
AAH71423
ID   AAH71423 standard; cDNA; 366 BP.
XX
AC   AAH71423;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 2697.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
```

PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US033312.
XX
PR 08-DEC-1999; 99US-0169681P.
PR 21-DEC-1999; 99US-0171350P.
PR 14-MAR-2000; 2000US-0189315P.
PR 12-MAY-2000; 2000US-0203791P.
PR 09-JUN-2000; 2000US-0210600P.
PR 21-JUL-2000; 2000US-0220114P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer and
PT for assessing and detecting compounds for treating the cancer.
XX
PS Claim 1; Page 555; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy
XX
SQ Sequence 366 BP; 98 A; 85 C; 76 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 28 AAGGTTATATAT 39

RESULT 25
AAH35822
ID AAH35822 standard; cDNA; 367 BP.
XX
AC AAH35822;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:2904.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG76417.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 4589-4590; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 367 BP; 113 A; 53 C; 61 G; 133 T; 0 U; 7 Other;

Query Match 100.0%; Score 12; DB 4; Length 367;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 133 AAGGTTATATAT 144

RESULT 26
AAS37872
ID AAS37872 standard; cDNA; 374 BP.
XX
AC AAS37872;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #930.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007787.
XX
PR 09-MAR-2000; 2000US-0188609P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and treatment
PT of breast, lung and colon cancer.
XX
PS Claim 1; Page 830-831; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for

CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention
XX
SQ Sequence 374 BP; 127 A; 72 C; 83 G; 92 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 196 AAGGTTATATAT 207

RESULT 27
ADV83971/c
ID ADV83971 standard; DNA; 375 BP.
XX
AC ADV83971;
XX
DT 24-FEB-2005 (first entry)
XX
DE Streptococcus agalactiae DNA sequence, SEQ ID 5112.
XX
KW Antibacterial; vaccine; bacterial infection; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200292818-A2.
XX
PD 21-NOV-2002.
XX
PF 26-APR-2002; 2002WO-IB003059.
XX
PR 26-APR-2001; 2001FR-00005642.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR WPI; 2004-101891/11.
XX
PT Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT and identification of therapeutic targets.
XX
PS Claim 4; SEQ ID NO 5112; 439pp; French.
XX
CC The present invention relates to novel Streptococcus agalactiae
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC synthesis of amino acids, cell membranes, intermediate (central)
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC regulatory functions, replication, transcription, translation, protein
CC transport, adaptation to atypical conditions, sensitivity to medicines
CC and/or analogues, functions related to transposons, biosynthesis of
CC cofactors, prosthetic groups and transporters, cell membrane proteins and
CC cellular machinery. (I) are useful for the detection and/or amplification
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC useful for treatment of a bacterial S. agalactiae infection. The complete
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC present patent is an equivalent for the basic patent FR2824074A1, which
CC contains only 2344 sequences.
XX
SQ Sequence 375 BP; 159 A; 44 C; 64 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 13; Length 375;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 164 AAGGTTATATAT 153

RESULT 28
AAK81895
ID AAK81895 standard; DNA; 378 BP.
XX
AC AAK81895;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36707.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 36707; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 378 BP; 119 A; 47 C; 76 G; 136 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 152 AAGGTTATATAT 163

RESULT 29
AAK81894
ID AAK81894 standard; DNA; 378 BP.
XX
AC AAK81894;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36706.

XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 36706; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 378 BP; 119 A; 47 C; 76 G; 136 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 152 AAGGTTATATAT 163

RESULT 30
AAK62195/c
ID AAK62195 standard; cDNA; 378 BP.
XX
AC AAK62195;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:7255.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

```
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM89414.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Claim 1; SEQ ID NO 7255; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 378 BP; 135 A; 76 C; 48 G; 116 T; 0 U; 3 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        227 AAGGTTATATAT 216


Search completed: December 20, 2005, 05:54:25
Job time : 12.2289 secs
```

THIS PAGE BLANK (USPTO)

-GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:12:49 ; Search time 2.68273 Seconds
(without alignments)
7951.129 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	12	3	US-08-098-327E-42	Sequence 42, Appl
	2	12	100.0	12	3	US-08-098-327E-46	Sequence 46, Appl
	3	12	100.0	12	3	US-08-462-625-42	Sequence 42, Appl
	4	12	100.0	12	3	US-08-462-625-46	Sequence 46, Appl
c	5	12	100.0	146	3	US-09-513-999C-23604	Sequence 23604, A
c	6	12	100.0	184	3	US-09-480-921B-13	Sequence 13, Appl
c	7	12	100.0	242	3	US-09-513-999C-9073	Sequence 9073, Ap
c	8	12	100.0	294	3	US-09-513-999C-22577	Sequence 22577, A
	9	12	100.0	362	3	US-09-513-999C-28463	Sequence 28463, A
c	10	12	100.0	502	3	US-10-002-623-775	Sequence 775, App
c	11	12	100.0	601	3	US-09-949-016-24319	Sequence 24319, A
c	12	12	100.0	601	3	US-09-949-016-24320	Sequence 24320, A
c	13	12	100.0	601	3	US-09-949-016-24321	Sequence 24321, A
c	14	12	100.0	601	3	US-09-949-016-25053	Sequence 25053, A
c	15	12	100.0	601	3	US-09-949-016-25054	Sequence 25054, A
	16	12	100.0	601	3	US-09-949-016-28868	Sequence 28868, A
	17	12	100.0	601	3	US-09-949-016-28869	Sequence 28869, A
c	18	12	100.0	601	3	US-09-949-016-32135	Sequence 32135, A
c	19	12	100.0	601	3	US-09-949-016-32136	Sequence 32136, A
c	20	12	100.0	601	3	US-09-949-016-32137	Sequence 32137, A
	21	12	100.0	601	3	US-09-949-016-32922	Sequence 32922, A
	22	12	100.0	601	3	US-09-949-016-32923	Sequence 32923, A
c	23	12	100.0	601	3	US-09-949-016-38188	Sequence 38188, A
c	24	12	100.0	601	3	US-09-949-016-60073	Sequence 60073, A
c	25	12	100.0	601	3	US-09-949-016-68392	Sequence 68392, A
c	26	12	100.0	601	3	US-09-949-016-74492	Sequence 74492, A
	27	12	100.0	601	3	US-09-949-016-76566	Sequence 76566, A
	28	12	100.0	601	3	US-09-949-016-80457	Sequence 80457, A
c	29	12	100.0	601	3	US-09-949-016-87731	Sequence 87731, A
c	30	12	100.0	601	3	US-09-949-016-94872	Sequence 94872, A
	31	12	100.0	601	3	US-09-949-016-115481	Sequence 115481,
	32	12	100.0	601	3	US-09-949-016-119463	Sequence 119463,
	33	12	100.0	601	3	US-09-949-016-119464	Sequence 119464,
c	34	12	100.0	601	3	US-09-949-016-122654	Sequence 122654,
c	35	12	100.0	601	3	US-09-949-016-125367	Sequence 125367,
c	36	12	100.0	601	3	US-09-949-016-125368	Sequence 125368,
	37	12	100.0	601	3	US-09-949-016-130126	Sequence 130126,
	38	12	100.0	601	3	US-09-949-016-130127	Sequence 130127,
	39	12	100.0	601	3	US-09-949-016-132207	Sequence 132207,
	40	12	100.0	601	3	US-09-949-016-132208	Sequence 132208,
	41	12	100.0	601	3	US-09-949-016-132275	Sequence 132275,
	42	12	100.0	601	3	US-09-949-016-153200	Sequence 153200,
	43	12	100.0	601	3	US-09-949-016-153201	Sequence 153201,
c	44	12	100.0	601	3	US-09-949-016-154936	Sequence 154936,
c	45	12	100.0	601	3	US-09-949-016-154937	Sequence 154937,

ALIGNMENTS

```
RESULT 1
US-08-098-327E-42
; Sequence 42, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
```

```
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
US-08-098-327E-46
; Sequence 46, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
US-08-462-625-42
; Sequence 42, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
US-08-462-625-46
; Sequence 46, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
```

```
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
US-09-513-999C-23604/c
; Sequence 23604, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 23604
;   LENGTH: 146
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-23604

  Query Match             100.0%;  Score 12;  DB 3;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 6
US-09-480-921B-13/c
; Sequence 13, Application US/09480921B
; Patent No. 6387637
; GENERAL INFORMATION:
;  APPLICANT: Levin, Joshua Z.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Potter, Sharon L.
;  APPLICANT:  Wegrich, Lynette M.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-30780A
;  CURRENT APPLICATION NUMBER: US/09/480,921B
;  CURRENT FILING DATE:  2000-01-11
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 184
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-480-921B-13

  Query Match             100.0%;  Score 12;  DB 3;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 7
US-09-513-999C-9073/c
; Sequence 9073, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 9073
;   LENGTH: 242
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 17
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: 28
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 37
;   OTHER INFORMATION: r=a or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 39
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 45
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 85
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 106
;   OTHER INFORMATION: b=c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 114
;   OTHER INFORMATION: w=a or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 156
;   OTHER INFORMATION: d=a or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 160
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 187
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 200
;   OTHER INFORMATION: m=a or c
US-09-513-999C-9073

  Query Match             100.0%;  Score 12;  DB 3;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 8
US-09-513-999C-22577/c
; Sequence 22577, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 22577
;   LENGTH: 294
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 33
;   OTHER INFORMATION: w=a or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 119
;   OTHER INFORMATION: m=a or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 195
;   OTHER INFORMATION: k=g or t
US-09-513-999C-22577

  Query Match             100.0%;  Score 12;  DB 3;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        240 AAGGTTATATAT 229


RESULT 9
US-09-513-999C-28463
; Sequence 28463, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 28463
;   LENGTH: 362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-28463

  Query Match             100.0%;  Score 12;  DB 3;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        295 AAGGTTATATAT 306


RESULT 10
US-10-002-623-775/c
; Sequence 775, Application US/10002623
; Patent No. 6929911
; GENERAL INFORMATION:
;  APPLICANT: OEFNER, PETER J.
;  APPLICANT:  UNDERHILL, PETER A.
;  TITLE OF INVENTION: A METHOD FOR DETERMINING GENETIC
;  TITLE OF INVENTION:  AFFILIATION, SUBSTRUCTURE AND GENE FLOW WITHIN HUMAN
;  TITLE OF INVENTION:  POPULATIONS
;  FILE REFERENCE: STAN-212
;  CURRENT APPLICATION NUMBER: US/10/002,623
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/245,355
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 952
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 775
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-002-623-775

Query Match 100.0%; Score 12; DB 3; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 83 AAGGTTATATAT 72

RESULT 11
US-09-949-016-24319/c
; Sequence 24319, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24319
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24319

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 595 AAGGTTATATAT 584

RESULT 12
US-09-949-016-24320/c
; Sequence 24320, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24320
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24320

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 557 AAGGTTATATAT 546

RESULT 13
US-09-949-016-24321/c
; Sequence 24321, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24321
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24321

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 23 AAGGTTATATAT 12

RESULT 14
US-09-949-016-25053/c
; Sequence 25053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25053
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-25053

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        564 AAGGTTATATAT 553


RESULT 15
US-09-949-016-25054/c
; Sequence 25054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25054
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-25054

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        216 AAGGTTATATAT 205


RESULT 16
US-09-949-016-28868
; Sequence 28868, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28868
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-28868

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        349 AAGGTTATATAT 360


RESULT 17
US-09-949-016-28869
; Sequence 28869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28869
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-28869

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        183 AAGGTTATATAT 194


RESULT 18
US-09-949-016-32135/c
; Sequence 32135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32135
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32135

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        358 AAGGTTATATAT 347


RESULT 19
US-09-949-016-32136/c
; Sequence 32136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32136
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32136

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        140 AAGGTTATATAT 129


RESULT 20
US-09-949-016-32137/c
; Sequence 32137, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32137
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32137

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         62 AAGGTTATATAT 51


RESULT 21
US-09-949-016-32922
; Sequence 32922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32922
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32922

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        244 AAGGTTATATAT 255


RESULT 22
US-09-949-016-32923
; Sequence 32923, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32923
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32923

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 23
US-09-949-016-38188/c
; Sequence 38188, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
```

```
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38188
;    LENGTH: 601
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-38188

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        575 AAGGTTATATAT 564


RESULT 24
US-09-949-016-60073/c
; Sequence 60073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60073
;    LENGTH: 601
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-60073

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        196 AAGGTTATATAT 185


RESULT 25
US-09-949-016-68392/c
; Sequence 68392, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68392
;    LENGTH: 601
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-68392

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        575 AAGGTTATATAT 564


RESULT 26
US-09-949-016-74492/c
; Sequence 74492, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74492
;    LENGTH: 601
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-74492

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        356 AAGGTTATATAT 345


RESULT 27
US-09-949-016-76566
; Sequence 76566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 76566
;    LENGTH: 601
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-76566
```

```
  Query Match           100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        339 AAGGTTATATAT 350


RESULT 28
US-09-949-016-80457
; Sequence 80457, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 80457
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-80457

  Query Match           100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        164 AAGGTTATATAT 175


RESULT 29
US-09-949-016-87731/c
; Sequence 87731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 87731
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-87731

  Query Match           100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        207 AAGGTTATATAT 196


RESULT 30
US-09-949-016-94872/c
; Sequence 94872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 94872
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-94872

  Query Match           100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        359 AAGGTTATATAT 348


Search completed: December 20, 2005, 15:19:08
Job time : 3.68273 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 10:43:25 ; Search time 13 Seconds
(without alignments)
7633.272 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	12	100.0	12	3	US-09-837-344-42	Sequence 42, Appl
	2	12	100.0	12	3	US-09-837-344-46	Sequence 46, Appl
	3	12	100.0	12	3	US-09-900-963-42	Sequence 42, Appl
	4	12	100.0	12	3	US-09-900-963-46	Sequence 46, Appl
	5	12	100.0	13	8	US-10-257-017B-193201	Sequence 193201,
c	6	12	100.0	13	8	US-10-257-017B-193202	Sequence 193202,
c	7	12	100.0	25	8	US-10-719-900-346370	Sequence 346370,
	8	12	100.0	100	6	US-10-029-386-22059	Sequence 22059, A
	9	12	100.0	146	6	US-10-029-386-14841	Sequence 14841, A
	10	12	100.0	176	3	US-09-969-373-199	Sequence 199, App
	11	12	100.0	176	3	US-09-969-373-200	Sequence 200, App
c	12	12	100.0	184	5	US-10-047-412A-13	Sequence 13, Appl
	13	12	100.0	188	3	US-09-754-853A-298	Sequence 298, App
c	14	12	100.0	190	8	US-10-674-124A-18049	Sequence 18049, A
c	15	12	100.0	201	8	US-10-719-993-30163	Sequence 30163, A
	16	12	100.0	201	8	US-10-741-600-19082	Sequence 19082, A
	17	12	100.0	201	8	US-10-741-600-45291	Sequence 45291, A
c	18	12	100.0	218	7	US-10-424-599-81702	Sequence 81702, A
c	19	12	100.0	232	8	US-10-674-124A-24402	Sequence 24402, A
c	20	12	100.0	242	8	US-10-425-115-54924	Sequence 54924, A
c	21	12	100.0	245	8	US-10-425-115-111652	Sequence 111652,
c	22	12	100.0	250	4	US-09-925-065A-650047	Sequence 650047,
	23	12	100.0	255	3	US-09-867-701-136	Sequence 136, App
	24	12	100.0	267	3	US-09-864-408A-4589	Sequence 4589, Ap
c	25	12	100.0	277	8	US-10-357-930-7109	Sequence 7109, Ap
	26	12	100.0	284	7	US-10-437-963-24548	Sequence 24548, A
c	27	12	100.0	289	3	US-09-867-701-64	Sequence 64, Appl
c	28	12	100.0	313	3	US-09-867-701-3452	Sequence 3452, Ap
c	29	12	100.0	313	8	US-10-674-124A-23482	Sequence 23482, A
	30	12	100.0	316	3	US-09-920-300A-1741	Sequence 1741, Ap
	31	12	100.0	316	5	US-10-033-528-1741	Sequence 1741, Ap
	32	12	100.0	316	6	US-10-099-926-1741	Sequence 1741, Ap
	33	12	100.0	316	9	US-10-961-527-1741	Sequence 1741, Ap
	34	12	100.0	322	8	US-10-425-115-138506	Sequence 138506,
	35	12	100.0	328	8	US-10-674-124A-13190	Sequence 13190, A
c	36	12	100.0	334	4	US-09-925-065A-115836	Sequence 115836,
c	37	12	100.0	334	7	US-10-242-535A-16001	Sequence 16001, A
c	38	12	100.0	334	7	US-10-085-783A-16001	Sequence 16001, A
c	39	12	100.0	337	7	US-10-242-535A-35808	Sequence 35808, A
c	40	12	100.0	337	7	US-10-085-783A-35808	Sequence 35808, A
	41	12	100.0	367	5	US-10-106-698-2914	Sequence 2914, Ap
c	42	12	100.0	373	8	US-10-425-115-27009	Sequence 27009, A
	43	12	100.0	374	3	US-09-803-719-930	Sequence 930, App
	44	12	100.0	374	9	US-10-779-543-14200	Sequence 14200, A
c	45	12	100.0	383	7	US-10-242-535A-17171	Sequence 17171, A

ALIGNMENTS

```
RESULT 1
US-09-837-344-42
; Sequence 42, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
```

```
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-837-344-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
US-09-837-344-46
; Sequence 46, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-837-344-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
US-09-900-963-42
; Sequence 42, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-900-963-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
US-09-900-963-46
```

```
; Sequence 46, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 12 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..12
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-900-963-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
US-10-257-017B-193201
; Sequence 193201, Application US/10257017B
; Publication No. US20040241651A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Detection of single nucleotide polymorhphisms [SNPs] and cytosine
;  TITLE OF INVENTION:  methylations
;  FILE REFERENCE: E01/1193/WO
;  CURRENT APPLICATION NUMBER: US/10/257,017B
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 382046
; SEQ ID NO 193201
;   LENGTH: 13
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide for detection of SNP TSC0047525
US-10-257-017B-193201

  Query Match             100.0%;  Score 12;  DB 8;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 6
US-10-257-017B-193202/c
; Sequence 193202, Application US/10257017B
; Publication No. US20040241651A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Detection of single nucleotide polymorhphisms [SNPs] and cytosine
;  TITLE OF INVENTION:  methylations
;  FILE REFERENCE: E01/1193/WO
;  CURRENT APPLICATION NUMBER: US/10/257,017B
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 382046
; SEQ ID NO 193202
;   LENGTH: 13
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide for detection of SNP TSC0047525
US-10-257-017B-193202

  Query Match             100.0%;  Score 12;  DB 8;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         13 AAGGTTATATAT 2


RESULT 7
US-10-719-900-346370/c
; Sequence 346370, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
;  APPLICANT: Xue Mei Zhou
;  TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
;  FILE REFERENCE: 3528.1
;  CURRENT APPLICATION NUMBER: US/10/719,900
;  CURRENT FILING DATE:  2003-11-20
;  PRIOR APPLICATION NUMBER: 60/427,808
;  PRIOR FILING DATE: 2002 11 20
;  NUMBER OF SEQ ID NOS: 982914
;  SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 346370
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-719-900-346370
```

```
  Query Match           100.0%;  Score 12;  DB 8;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         20 AAGGTTATATAT 9


RESULT 8
US-10-029-386-22059
; Sequence 22059, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22059
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR14.3
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
;   OTHER INFORMATION: NT HIT: AB000381.1, EVALUE 8.00e-02
;   OTHER INFORMATION: EST_HUMAN HIT: H08306.1, EVALUE 4.70e-01
;   OTHER INFORMATION: SWISSPROT HIT: P30796, EVALUE 6.10e+00
US-10-029-386-22059

  Query Match           100.0%;  Score 12;  DB 6;  Length 100;
  Best Local Similarity  100.0%;  Pred. No. 9.9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         44 AAGGTTATATAT 55


RESULT 9
US-10-029-386-14841
; Sequence 14841, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14841
;   LENGTH: 146
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL121957.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.57
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.45
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.44
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.42
;   OTHER INFORMATION: EST_HUMAN HIT: BG958104.1, EVALUE 7.30e-01
;   OTHER INFORMATION: NT HIT: gi15295339, EVALUE 5.00e-01
US-10-029-386-14841

  Query Match           100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity  100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         97 AAGGTTATATAT 108


RESULT 10
US-09-969-373-199
; Sequence 199, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 199
;   LENGTH: 176
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-199

  Query Match           100.0%;  Score 12;  DB 3;  Length 176;
  Best Local Similarity  100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         21 AAGGTTATATAT 32


RESULT 11
US-09-969-373-200
; Sequence 200, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 200
;   LENGTH: 176
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-200

  Query Match           100.0%;  Score 12;  DB 3;  Length 176;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         21 AAGGTTATATAT 32


RESULT 12
US-10-047-412A-13/c
; Sequence 13, Application US/10047412A
; Publication No. US20020197696A1
; GENERAL INFORMATION:
;  APPLICANT: Levin, Joshua Z.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Potter, Sharon L.
;  APPLICANT:  Wegrich, Lynette M.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-30780DIV
;  CURRENT APPLICATION NUMBER: US/10/047,412A
;  CURRENT FILING DATE:  2002-04-11
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 184
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-10-047-412A-13

  Query Match             100.0%;  Score 12;  DB 5;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 13
US-09-754-853A-298
; Sequence 298, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 298
;   LENGTH: 188
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 318O13_region_A3__153114_12
US-09-754-853A-298

  Query Match             100.0%;  Score 12;  DB 3;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 14
US-10-674-124A-18049/c
; Sequence 18049, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 18049
;   LENGTH: 190
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: D11S1340
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 11
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 123440912
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 149287
US-10-674-124A-18049

  Query Match             100.0%;  Score 12;  DB 8;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        164 AAGGTTATATAT 153


RESULT 15
US-10-719-993-30163/c
; Sequence 30163, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30163
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-30163

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
```

```
              ||||||||||||
Db         63 AAGGTTATATAT 52


RESULT 16
US-10-741-600-19082
; Sequence 19082, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19082
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-19082

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         38 AAGGTTATATAT 49


RESULT 17
US-10-741-600-45291
; Sequence 45291, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45291
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-45291

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         73 AAGGTTATATAT 84


RESULT 18
US-10-424-599-81702/c
; Sequence 81702, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 81702
;   LENGTH: 218
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(218)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_44794C.1
US-10-424-599-81702

  Query Match             100.0%;  Score 12;  DB 7;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         28 AAGGTTATATAT 17


RESULT 19
US-10-674-124A-24402/c
; Sequence 24402, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 24402
;   LENGTH: 232
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: chr20.fa.07frz.6706738
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 20
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 6258865
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 90016
US-10-674-124A-24402

  Query Match             100.0%;  Score 12;  DB 8;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 85
```

```
RESULT 20
US-10-425-115-54924/c
; Sequence 54924, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 54924
;   LENGTH: 242
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(242)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_150089C.1
US-10-425-115-54924

  Query Match           100.0%;  Score 12;  DB 8;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        187 AAGGTTATATAT 176


RESULT 21
US-10-425-115-111652/c
; Sequence 111652, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 111652
;   LENGTH: 245
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_33313C.1
US-10-425-115-111652

  Query Match           100.0%;  Score 12;  DB 8;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        226 AAGGTTATATAT 215


RESULT 22
US-09-925-065A-650047/c
; Sequence 650047, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 650047
;   LENGTH: 250
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-650047

  Query Match           100.0%;  Score 12;  DB 4;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         51 AAGGTTATATAT 40


RESULT 23
US-09-867-701-136
; Sequence 136, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 136
;   LENGTH: 255
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-136

  Query Match           100.0%;  Score 12;  DB 3;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37


RESULT 24
US-09-864-408A-4589
; Sequence 4589, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Shimkets, Richard A.
;  TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Enco
;  FILE REFERENCE: 21402-012
;  CURRENT APPLICATION NUMBER: US/09/864,408A
```

```
;  CURRENT FILING DATE:  2001-05-24
;  PRIOR APPLICATION NUMBER: 60/206,690
;  PRIOR FILING DATE: 2000-05-24
;  NUMBER OF SEQ ID NOS: 9068
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4589
;   LENGTH: 267
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13)..(13)
;   OTHER INFORMATION: Wherein n may be a, c, g or t
US-09-864-408A-4589

  Query Match            100.0%;  Score 12;  DB 3;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         23 AAGGTTATATAT 34


RESULT 25
US-10-357-930-7109/c
; Sequence 7109, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007BCN
;  CURRENT APPLICATION NUMBER: US/10/357,930
;  CURRENT FILING DATE:  2003-02-04
;  PRIOR APPLICATION NUMBER: 09/785,276
;  PRIOR FILING DATE: 2003-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7109
;   LENGTH: 277
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 90, 257
;   OTHER INFORMATION: n = A,T,C or G
US-10-357-930-7109

  Query Match            100.0%;  Score 12;  DB 8;  Length 277;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         76 AAGGTTATATAT 65


RESULT 26
US-10-437-963-24548
; Sequence 24548, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24548
;   LENGTH: 284
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_29521C.1
US-10-437-963-24548

  Query Match            100.0%;  Score 12;  DB 7;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        178 AAGGTTATATAT 189


RESULT 27
US-09-867-701-64/c
; Sequence 64, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 64
;   LENGTH: 289
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-867-701-64

  Query Match            100.0%;  Score 12;  DB 3;  Length 289;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        250 AAGGTTATATAT 239


RESULT 28
US-09-867-701-3452/c
; Sequence 3452, Application US/09867701
```

; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3452
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(313)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-3452

Query Match 100.0%; Score 12; DB 3; Length 313;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 303 AAGGTTATATAT 292

RESULT 29
US-10-674-124A-23482/c
; Sequence 23482, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hidetoshi
; APPLICANT: TAMIYA, Gen
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; TITLE OF INVENTION: GENETIC POLYMORPHISM MARKERS
; FILE REFERENCE: ORIN-003CIP
; CURRENT APPLICATION NUMBER: US/10/674,124A
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 23482
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: AC067801.2_170716
; FEATURE:
; OTHER INFORMATION: Located on chromosome 18
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; OTHER INFORMATION: sequence : 43376825
; FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 137159
US-10-674-124A-23482

Query Match 100.0%; Score 12; DB 8; Length 313;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 224 AAGGTTATATAT 213

RESULT 30
US-09-920-300A-1741
; Sequence 1741, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1741
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13, 22, 237, 298
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-1741

Query Match 100.0%; Score 12; DB 3; Length 316;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 161 AAGGTTATATAT 172

Search completed: December 20, 2005, 23:01:59
Job time : 14 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 15:08:19 ; Search time 3.30924 Seconds
(without alignments)
1777.494 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161431 seqs, 245089505 residues

Total number of hits satisfying chosen parameters: 8322862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	201	6	US-10-995-561-14708	Sequence 14708, A
c	2	12	100.0	600	6	US-10-750-185-1498	Sequence 1498, Ap
c	3	12	100.0	605	6	US-10-750-185-658	Sequence 658, App
c	4	12	100.0	802	6	US-10-750-185-33511	Sequence 33511, A
	5	12	100.0	963	6	US-10-750-185-28663	Sequence 28663, A
c	6	12	100.0	1026	6	US-10-750-185-33249	Sequence 33249, A
c	7	12	100.0	1174	6	US-10-750-185-52955	Sequence 52955, A
	8	12	100.0	1327	6	US-10-750-185-41599	Sequence 41599, A
	9	12	100.0	1332	6	US-10-750-185-62132	Sequence 62132, A
c	10	12	100.0	1404	6	US-10-750-185-26764	Sequence 26764, A
	11	12	100.0	1439	6	US-10-750-185-52757	Sequence 52757, A
	12	12	100.0	1650	6	US-10-750-185-54339	Sequence 54339, A
c	13	12	100.0	1913	6	US-10-750-185-27237	Sequence 27237, A
	14	12	100.0	2004	6	US-10-750-185-58707	Sequence 58707, A
	15	12	100.0	2050	6	US-10-750-185-37205	Sequence 37205, A
c	16	12	100.0	2269	7	US-11-087-072-1	Sequence 1, Appli
	17	12	100.0	2474	6	US-10-750-185-31559	Sequence 31559, A
c	18	12	100.0	2685	6	US-10-750-185-39093	Sequence 39093, A
	19	12	100.0	2688	6	US-10-750-185-55769	Sequence 55769, A
	20	12	100.0	3118	6	US-10-793-626-3939	Sequence 3939, Ap
	21	12	100.0	3187	6	US-10-793-626-4381	Sequence 4381, Ap
c	22	12	100.0	3320	6	US-10-909-125-797	Sequence 797, App
	23	12	100.0	3712	6	US-10-750-185-61739	Sequence 61739, A
c	24	12	100.0	4880	6	US-10-750-185-31043	Sequence 31043, A
c	25	12	100.0	6192	6	US-10-750-185-44297	Sequence 44297, A
	26	12	100.0	13363	6	US-10-995-561-13446	Sequence 13446, A
	27	12	100.0	29959	6	US-10-995-561-13475	Sequence 13475, A
c	28	12	100.0	33019	6	US-10-995-561-13371	Sequence 13371, A
c	29	12	100.0	57917	6	US-10-995-561-13299	Sequence 13299, A
	30	12	100.0	95832	6	US-10-995-561-13273	Sequence 13273, A
c	31	12	100.0	103931	7	US-11-117-187-193	Sequence 193, App
c	32	12	100.0	149419	7	US-11-112-908-49	Sequence 49, Appl
c	33	12	100.0	161726	7	US-11-112-908-48	Sequence 48, Appl
c	34	12	100.0	161726	7	US-11-112-908-52	Sequence 52, Appl
c	35	12	100.0	166111	7	US-11-112-908-47	Sequence 47, Appl
	36	12	100.0	180654	7	US-11-121-086-58	Sequence 58, Appl
c	37	12	100.0	185393	7	US-11-121-086-101	Sequence 101, App
c	38	12	100.0	187745	7	US-11-121-086-83	Sequence 83, Appl
c	39	12	100.0	188682	7	US-11-112-908-23	Sequence 23, Appl
c	40	12	100.0	191797	7	US-11-121-086-13	Sequence 13, Appl
	41	12	100.0	209822	6	US-10-995-561-13198	Sequence 13198, A
	42	12	100.0	212716	7	US-11-121-086-95	Sequence 95, Appl
c	43	12	100.0	212716	7	US-11-121-086-95	Sequence 95, Appl
c	44	12	100.0	319608	7	US-11-145-703-1	Sequence 1, Appli
	45	12	100.0	387780	6	US-10-995-561-13259	Sequence 13259, A

ALIGNMENTS

RESULT 1
US-10-995-561-14708
; Sequence 14708, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14708
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-14708

Query Match 100.0%; Score 12; DB 6; Length 201;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         38 AAGGTTATATAT 49
```

RESULT 2
US-10-750-185-1498/c
; Sequence 1498, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482

```
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 1498
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bovine  MMBT17112
US-10-750-185-1498

  Query Match             100.0%;  Score 12;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        553 AAGGTTATATAT 542


RESULT 3
US-10-750-185-658/c
; Sequence 658, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 658
;   LENGTH: 605
;   TYPE: DNA
;   ORGANISM: Bovine  MMBT03522
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(5)
;   OTHER INFORMATION: n is any nucleotide
US-10-750-185-658

  Query Match             100.0%;  Score 12;  DB 6;  Length 605;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        345 AAGGTTATATAT 334


RESULT 4
US-10-750-185-33511/c
; Sequence 33511, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33511
;   LENGTH: 802
;   TYPE: DNA
;   ORGANISM: Bovine  19866880419621
US-10-750-185-33511

  Query Match             100.0%;  Score 12;  DB 6;  Length 802;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        631 AAGGTTATATAT 620


RESULT 5
US-10-750-185-28663
; Sequence 28663, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 28663
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Bovine  19866880781472
US-10-750-185-28663

  Query Match             100.0%;  Score 12;  DB 6;  Length 963;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         54 AAGGTTATATAT 65


RESULT 6
US-10-750-185-33249/c
; Sequence 33249, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
```

```
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33249
;   LENGTH: 1026
;   TYPE: DNA
;   ORGANISM: Bovine  19866881769644
US-10-750-185-33249

  Query Match             100.0%;  Score 12;  DB 6;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        162 AAGGTTATATAT 151


RESULT 7
US-10-750-185-52955/c
; Sequence 52955, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52955
;   LENGTH: 1174
;   TYPE: DNA
;   ORGANISM: Bovine  19866881042220
US-10-750-185-52955

  Query Match             100.0%;  Score 12;  DB 6;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         29 AAGGTTATATAT 18


RESULT 8
US-10-750-185-41599
; Sequence 41599, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41599
;   LENGTH: 1327
;   TYPE: DNA
;   ORGANISM: Bovine  19866881435791
US-10-750-185-41599

  Query Match             100.0%;  Score 12;  DB 6;  Length 1327;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        944 AAGGTTATATAT 955


RESULT 9
US-10-750-185-62132
; Sequence 62132, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 62132
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Bovine  19866880526631
US-10-750-185-62132

  Query Match             100.0%;  Score 12;  DB 6;  Length 1332;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        411 AAGGTTATATAT 422


RESULT 10
US-10-750-185-26764/c
; Sequence 26764, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 26764
;   LENGTH: 1404
;   TYPE: DNA
;   ORGANISM: Bovine  19866881239706
US-10-750-185-26764
```

```
  Query Match             100.0%;  Score 12;  DB 6;  Length 1404;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        965 AAGGTTATATAT 954


RESULT 11
US-10-750-185-52757
; Sequence 52757, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52757
;   LENGTH: 1439
;   TYPE: DNA
;   ORGANISM: Bovine  19866880951582
US-10-750-185-52757

  Query Match             100.0%;  Score 12;  DB 6;  Length 1439;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        455 AAGGTTATATAT 466


RESULT 12
US-10-750-185-54339
; Sequence 54339, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 54339
;   LENGTH: 1650
;   TYPE: DNA
;   ORGANISM: Bovine  19866880458349
US-10-750-185-54339

  Query Match             100.0%;  Score 12;  DB 6;  Length 1650;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        341 AAGGTTATATAT 352


RESULT 13
US-10-750-185-27237/c
; Sequence 27237, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27237
;   LENGTH: 1913
;   TYPE: DNA
;   ORGANISM: Bovine  19866880640589
US-10-750-185-27237

  Query Match             100.0%;  Score 12;  DB 6;  Length 1913;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        338 AAGGTTATATAT 327


RESULT 14
US-10-750-185-58707
; Sequence 58707, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 58707
;   LENGTH: 2004
;   TYPE: DNA
;   ORGANISM: Bovine  19866880608793
US-10-750-185-58707

  Query Match             100.0%;  Score 12;  DB 6;  Length 2004;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
```

```
                  ||||||||||||
Db        793 AAGGTTATATAT 804


RESULT 15
US-10-750-185-37205
; Sequence 37205, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 37205
;   LENGTH: 2050
;   TYPE: DNA
;   ORGANISM: Bovine  19866880646640
US-10-750-185-37205

  Query Match             100.0%;  Score 12;  DB 6;  Length 2050;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1771 AAGGTTATATAT 1782


RESULT 16
US-11-087-072-1/c
; Sequence 1, Application US/11087072
; Publication No. US20050272069A1
;   GENERAL INFORMATION:
;        APPLICANT: Olopade, Olufunmilayo I.
;        TITLE OF INVENTION: METHYLTHIOADENOSINE PHOSPHORYLASE
;                            COMPOSITIONS AND METHODS OF USE IN
;                            THE DIAGNOSIS AND TREATMENT OF
;                            PROLIFERATIVE DISORDERS
;        NUMBER OF SEQUENCES: 15
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: United States of America
;             ZIP: 77210
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/11/087,072
;             FILING DATE: 22-Mar-2005
;             CLASSIFICATION: 435
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/674,311
;             FILING DATE: 01-JUL-1996
;             APPLICATION NUMBER: US 60/000,831
;             FILING DATE: 02-JUL-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kitchell, Barbara S.
;             REGISTRATION NUMBER: 33,928
;             REFERENCE/DOCKET NUMBER: ARSB:509
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2269 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  122..970
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-11-087-072-1

  Query Match             100.0%;  Score 12;  DB 7;  Length 2269;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1776 AAGGTTATATAT 1765


RESULT 17
US-10-750-185-31559
; Sequence 31559, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31559
;   LENGTH: 2474
;   TYPE: DNA
;   ORGANISM: Bovine  19866880685931
US-10-750-185-31559

  Query Match             100.0%;  Score 12;  DB 6;  Length 2474;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1439 AAGGTTATATAT 1450


RESULT 18
US-10-750-185-39093/c
; Sequence 39093, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
```

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39093
; LENGTH: 2685
; TYPE: DNA
; ORGANISM: Bovine 19866880416817
US-10-750-185-39093

Query Match 100.0%; Score 12; DB 6; Length 2685;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 223 AAGGTTATATAT 212

RESULT 19
US-10-750-185-55769
; Sequence 55769, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55769
; LENGTH: 2688
; TYPE: DNA
; ORGANISM: Bovine 19866880540328
US-10-750-185-55769

Query Match 100.0%; Score 12; DB 6; Length 2688;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 2350 AAGGTTATATAT 2361

RESULT 20
US-10-793-626-3939
; Sequence 3939, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 3118
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3939

Query Match 100.0%; Score 12; DB 6; Length 3118;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 1803 AAGGTTATATAT 1814

RESULT 21
US-10-793-626-4381
; Sequence 4381, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4381
; LENGTH: 3187
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4381

Query Match 100.0%; Score 12; DB 6; Length 3187;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 593 AAGGTTATATAT 604

RESULT 22
US-10-909-125-797/c
; Sequence 797, Application US/10909125
; Publication No. US20050261218A1
; GENERAL INFORMATION:
; APPLICANT: Esau, Christine
; APPLICANT: Lollo, Bridget
; APPLICANT: Bennett, C. Frank
; APPLICANT: Freier, Susan M.
; APPLICANT: Griffey, Richard H.
; APPLICANT: Baker, Brenda F.
; APPLICANT: Vickers, Timothy
; APPLICANT: Marcusson, Eric G.
; APPLICANT: Koller, Erich
; APPLICANT: Swayze, Eric
; APPLICANT: Jain, Ravi
; APPLICANT: Bhat, Balkrishen
; APPLICANT: Peralta, Eigen
; TITLE OF INVENTION: Oligomeric Compounds And Compositions For Use In Modulation
; TITLE OF INVENTION: Of Small Non-Coding RNAs
; FILE REFERENCE: ISIS0080-100 (CORE0016US)
; CURRENT APPLICATION NUMBER: US/10/909,125
; CURRENT FILING DATE: 2004-07-30
; PRIOR APPLICATION NUMBER: US 60/492,056

```
;  PRIOR FILING DATE: 2003-07-31
;  PRIOR APPLICATION NUMBER: US 60/516,303
;  PRIOR FILING DATE: 2003-10-31
;  PRIOR APPLICATION NUMBER: US 60/531,596
;  PRIOR FILING DATE: 2003-12-19
;  PRIOR APPLICATION NUMBER: US 60/562,417
;  PRIOR FILING DATE: 2004-04-14
;  NUMBER OF SEQ ID NOS: 2184
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 797
;   LENGTH: 3320
;   TYPE: DNA
;   ORGANISM: H. sapiens
US-10-909-125-797

  Query Match             100.0%;  Score 12;  DB 6;  Length 3320;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       2662 AAGGTTATATAT 2651


RESULT 23
US-10-750-185-61739
; Sequence 61739, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 61739
;   LENGTH: 3712
;   TYPE: DNA
;   ORGANISM: Bovine  19866881126830
US-10-750-185-61739

  Query Match             100.0%;  Score 12;  DB 6;  Length 3712;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       2143 AAGGTTATATAT 2154


RESULT 24
US-10-750-185-31043/c
; Sequence 31043, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31043
;   LENGTH: 4880
;   TYPE: DNA
;   ORGANISM: Bovine  19866880412972
US-10-750-185-31043

  Query Match             100.0%;  Score 12;  DB 6;  Length 4880;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       4367 AAGGTTATATAT 4356


RESULT 25
US-10-750-185-44297/c
; Sequence 44297, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 44297
;   LENGTH: 6192
;   TYPE: DNA
;   ORGANISM: Bovine  19866881113731
US-10-750-185-44297

  Query Match             100.0%;  Score 12;  DB 6;  Length 6192;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        665 AAGGTTATATAT 654


RESULT 26
US-10-995-561-13446
; Sequence 13446, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13446
;   LENGTH: 13363
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-10-995-561-13446

  Query Match             100.0%;  Score 12;  DB 6;  Length 13363;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      13052 AAGGTTATATAT 13063


RESULT 27
US-10-995-561-13475
; Sequence 13475, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13475
;   LENGTH: 29959
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13475

  Query Match             100.0%;  Score 12;  DB 6;  Length 29959;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      19791 AAGGTTATATAT 19802


RESULT 28
US-10-995-561-13371/c
; Sequence 13371, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13371
;   LENGTH: 33019
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13371

  Query Match             100.0%;  Score 12;  DB 6;  Length 33019;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      13394 AAGGTTATATAT 13383


RESULT 29
US-10-995-561-13299/c
; Sequence 13299, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13299
;   LENGTH: 57917
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13299

  Query Match             100.0%;  Score 12;  DB 6;  Length 57917;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      55957 AAGGTTATATAT 55946


RESULT 30
US-10-995-561-13273
; Sequence 13273, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13273
;   LENGTH: 95832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13273

  Query Match             100.0%;  Score 12;  DB 6;  Length 95832;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      76948 AAGGTTATATAT 76959


Search completed: December 20, 2005, 23:15:52
Job time : 7.30924 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:09:34 ; Search time 63.7751 Seconds
(without alignments)
8803.514 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		12	100.0	66	9	CC884412	CC884412 SALK_1129
2		12	100.0	73	10	AL757665	AL757665 Arabidops
3		12	100.0	94	10	BX897322	BX897322 Arabidops
c 4		12	100.0	99	3	BJ031417	BJ031417 BJ031417
c 5		12	100.0	102	1	AW066376	AW066376 687001E09
6		12	100.0	102	5	BW423536	BW423536 BW423536
c 7		12	100.0	106	1	AI943870	AI943870 614030E12
c 8		12	100.0	109	3	BI847701	BI847701 469886 MA
c 9		12	100.0	112	1	AI810434	AI810434 wb84f01.x
c 10		12	100.0	116	9	CC428121	CC428121 PUHBC05TB
c 11		12	100.0	116	10	BX535978	BX535978 Arabidops
c 12		12	100.0	117	10	BX535977	BX535977 Arabidops
c 13		12	100.0	135	10	CL883998	CL883998 abf64a08.
14		12	100.0	137	7	CK923106	CK923106 p3fmgc_01
15		12	100.0	146	5	BX327794	BX327794 BX327794
c 16		12	100.0	147	9	B95930	B95930 T18I6TF TAM
c 17		12	100.0	148	6	CA364088	CA364088 638929 NC
18		12	100.0	150	2	BG929548	BG929548 950006A12
19		12	100.0	153	1	AV107977	AV107977 AV107977
c 20		12	100.0	155	2	BE058406	BE058406 sn15f09.y
21		12	100.0	157	6	CA766087	CA766087 AF53-Rpf_
c 22		12	100.0	158	7	CO720276	CO720276 tai41f12.
c 23		12	100.0	159	8	T26907	T26907 ESTDIR575 C
24		12	100.0	160	6	CD162149	CD162149 ML1-0075T
c 25		12	100.0	161	2	BG513953	BG513953 dac15h11.
26		12	100.0	163	10	CE584959	CE584959 tigr-gss-
c 27		12	100.0	165	10	BX655471	BX655471 Arabidops
c 28		12	100.0	166	7	CV314160	CV314160 CM1-FN010
29		12	100.0	168	2	BF911105	BF911105 PM1-UT006
c 30		12	100.0	168	3	BP616314	BP616314 BP616314
31		12	100.0	168	10	CE297732	CE297732 tigr-gss-
32		12	100.0	169	9	BH559793	BH559793 BOGVD49TR
33		12	100.0	170	1	AV106575	AV106575 AV106575
34		12	100.0	172	9	BH532142	BH532142 BOGVD38TR
35		12	100.0	172	10	CE853534	CE853534 tigr-gss-
36		12	100.0	173	2	BG937368	BG937368 950006A12
37		12	100.0	174	1	AV106563	AV106563 AV106563
38		12	100.0	177	1	BB000380	BB000380 BB000380
c 39		12	100.0	177	3	BP618736	BP618736 BP618736
40		12	100.0	178	1	AV104185	AV104185 AV104185
c 41		12	100.0	178	6	CD024554	CD024554 NXRV_044_
42		12	100.0	178	7	CV385592	CV385592 QV1-BT082
43		12	100.0	180	10	CW801296	CW801296 WiscDsLox
44		12	100.0	186	1	AI842967	AI842967 UI-M-AO1-
c 45		12	100.0	187	9	AZ830376	AZ830376 2M0109P06

ALIGNMENTS

RESULT 1
CC884412
LOCUS CC884412 66 bp DNA linear GSS 31-JUL-2003
DEFINITION SALK_112931.20.95.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_112931.20.95.x, genomic survey sequence.
ACCESSION CC884412
VERSION CC884412.1 GI:33360768
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 66)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At3g13660.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .66
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_112931.20.95.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 27 AAGGTTATATAT 38

RESULT 2
AL757665
LOCUS AL757665 73 bp DNA linear GSS 01-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-151D09-013133, genomic survey sequence.
ACCESSION AL757665
VERSION AL757665.1 GI:21496013
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 73)
AUTHORS Rosso,M.G., Li,Y., Strizhov,N. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At2g29840. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .73
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-151D09-013133"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion.

T-DNA derived sequences were removed."
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 73;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 46 AAGGTTATATAT 57

RESULT 3
BX897322
LOCUS BX897322 94 bp DNA linear GSS 05-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-766A12-023593, genomic survey sequence.
ACCESSION BX897322
VERSION BX897322.1 GI:39929817
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 94)
AUTHORS Strizhov,N., Rosso,M.G., Li,Y. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At1g53030. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .94
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-766A12-023593"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion.

T-DNA derived sequences were removed."
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 69 AAGGTTATATAT 80

RESULT 4
BJ031417/c
LOCUS BJ031417 99 bp mRNA linear EST 26-SEP-2003
DEFINITION BJ031417 NIBB Mochii normalized Xenopus neurula library Xenopus laevis cDNA clone XL007b24 5', mRNA sequence.
ACCESSION BJ031417
VERSION BJ031417.1 GI:17408235
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 99)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .99
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL007b24"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/clone_lib="NIBB Mochii normalized Xenopus neurula library"
ORIGIN

Query Match 100.0%; Score 12; DB 3; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 27 AAGGTTATATAT 16

RESULT 5
AW066376/c
LOCUS AW066376 102 bp mRNA linear EST 30-MAR-2000
DEFINITION 687001E09.y1 687 - Early embryo from Delaware Zea mays cDNA, mRNA sequence.
ACCESSION AW066376
VERSION AW066376.1 GI:6021448
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 102)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687001 row: E column: 09.
FEATURES Location/Qualifiers
source 1. .102
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/tissue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/clone_lib="687 - Early embryo from Delaware"
/note="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI; Site_2: EcoRI; Library was prepared by Stratagene using the Uni-ZAP XR system (Stratagene BN937328-12). Clones were picked by a Q-bot after blue/white selection (ampicillin resistance - use 100 micrograms/microliter). Developed from a pool of equal amounts of RNA from developing embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize Strain Cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"
ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 97 AAGGTTATATAT 86

RESULT 6
BW423536
LOCUS BW423536 102 bp mRNA linear EST 02-JUN-2004
DEFINITION BW423536 Yutaka Satou unpublished cDNA library, mature adult whole animal Ciona intestinalis cDNA clone cima846f02 3', mRNA sequence.
ACCESSION BW423536
VERSION BW423536.1 GI:47961086
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 102)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Yutaka Satou
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES Location/Qualifiers
source 1. .102
/organism="Ciona intestinalis"
/mol_type="mRNA"

```
                     /db_xref="taxon:7719"
                     /clone="cima846f02"
                     /tissue_type="whole animal"
                     /dev_stage="mature adult"
                     /clone_lib="Yutaka Satou unpublished cDNA library, mature
                     adult whole animal"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 5;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         15 AAGGTTATATAT 26


RESULT 7
AI943870/c
LOCUS       AI943870                 106 bp    mRNA    linear   EST 09-AUG-1999
DEFINITION  614030E12.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI943870
VERSION     AI943870.1  GI:5713885
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 106)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614030  row: E  column: 12.
FEATURES             Location/Qualifiers
     source          1. .106
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 1;  Length 106;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        104 AAGGTTATATAT 93


RESULT 8
BI847701/c
LOCUS       BI847701                 109 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  469886 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI847701
VERSION     BI847701.1  GI:15960220
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: L  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .109
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 3;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         70 AAGGTTATATAT 59


RESULT 9
AI810434/c
LOCUS       AI810434                 112 bp    mRNA    linear   EST 17-DEC-1999
DEFINITION  wb84f01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2312377 3',
            mRNA sequence.
ACCESSION   AI810434
VERSION     AI810434.1  GI:5397000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 112)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 469 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 79.

FEATURES Location/Qualifiers
source 1. .112
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2312377"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Lb 41 AAGGTTATATAT 30

RESULT 10
CC428121/c
LOCUS CC428121 116 bp DNA linear GSS 19-MAY-2003
DEFINITION PUHBC05TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa411A09, genomic survey sequence.
ACCESSION CC428121
VERSION CC428121.1 GI:30908211
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 116)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES Location/Qualifiers
source 1. .116
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa411A09"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"

ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 115 AAGGTTATATAT 104

RESULT 11
BX535978/c
LOCUS BX535978 116 bp DNA linear GSS 04-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-522G02-020204, genomic survey sequence.
ACCESSION BX535978
VERSION BX535978.1 GI:31413108
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 116)
AUTHORS Strizhov,N., Rosso,M.G., Li,Y. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At5g54040. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES Location/Qualifiers
source 1. .116
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-522G02-020204"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from

vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        111 AAGGTTATATAT 100
```

RESULT 12
BX535977/c
LOCUS BX535977 117 bp DNA linear GSS 04-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-522G02-020049, genomic survey sequence.
ACCESSION BX535977
VERSION BX535977.1 GI:31413107
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 117)
AUTHORS Li,Y., Rosso,M.G., Strizhov,N. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At5g54040. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .117
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-522G02-020049"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from

vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        112 AAGGTTATATAT 101
```

RESULT 13
CL883998/c
LOCUS CL883998 135 bp DNA linear GSS 30-AUG-2004
DEFINITION abf64a08.y1 Soybean random, unfiltered genomic library Glycine max genomic, genomic survey sequence.
ACCESSION CL883998
VERSION CL883998.1 GI:51622431
KEYWORDS GSS.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 135)
AUTHORS Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J., Peterson,S., Jones,J., Fries,J., Budiman,M.A., Nguyen,H. and Stacey,G.
TITLE Methylation filtered genomic sequences from Glycine max
JOURNAL Unpublished (2004)
COMMENT Contact: Gary Stacey
University of Missouri
108 Waters Hall, Columbia, MO 65211, USA
Tel: 573-884-1267
Fax: 573-882-0588
Email: staceyg@missouri.edu
LidID: 230
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .135
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Williams 82"
/db_xref="taxon:3847"
/tissue_type="Young leaves"
/clone_lib="Soybean random, unfiltered genomic library"
/note="Vector: pOT2; Site_1: BstXI; Randomly sheared genomic DNA ranging from 0.7-1.5 kb were end repaired and ligated to BstXI linkers prior to cloning in BstXI-cut pOT2. LibID: 230"
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         18 AAGGTTATATAT 7
```

RESULT 14
CK923106
LOCUS CK923106 137 bp mRNA linear EST 22-APR-2004
DEFINITION p3fmgc_012733 Normalized Magnaporthe grisea cDNA pRS423 library Magnaporthe grisea cDNA clone p3fmgc_012733, mRNA sequence.
ACCESSION CK923106

```
VERSION     CK923106.1  GI:45391458
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Dong,H., Peng,Y., Chen,B., Li,Y. and Li,D.
  TITLE     Large-scale identification of ESTs from Magnaporthe grisea by
            normalized cDNA library sequencing (2004b)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University, China Agricultural University, Guangxi
            University
            Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, china
            Tel: 0086-571-86961525
            Fax: 0086-571-86971183
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 reverse primer.
FEATURES             Location/Qualifiers
     source          1. .137
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /db_xref="taxon:148305"
                     /clone="p3fmgc_012733"
                     /tissue_type="Mycellium, conidium, germinating conidium,
                     swelling appressorium, mature appressorium, penetration
                     peg"
                     /dev_stage="Mycellium, conidium, germinating conidium,
                     swelling appressorium, mature appressorium, penetration
                     peg"
                     /clone_lib="Normalized Magnaporthe grisea cDNA pRS423
                     library"
                     /note="Vector: pRS423"
ORIGIN

  Query Match            100.0%;  Score 12;  DB 7;  Length 137;
  Best Local Similarity  100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        126 AAGGTTATATAT 137


RESULT 15
BX327794
LOCUS       BX327794               146 bp    mRNA    linear   EST 07-APR-2004
DEFINITION  BX327794 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI026YN16 5-PRIME, mRNA sequence.
ACCESSION   BX327794
VERSION     BX327794.2  GI:46270728
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 2, 2003 this sequence version replaced gi:30336692.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            1798.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAG028ZH05_CS02632_1&c=1798.r
            .
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI026YN16"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match            100.0%;  Score 12;  DB 5;  Length 146;
  Best Local Similarity  100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          6 AAGGTTATATAT 17


RESULT 16
B95930/c
LOCUS       B95930                 147 bp    DNA     linear   GSS 31-MAR-1998
DEFINITION  T18I6TF TAMU Arabidopsis thaliana genomic clone T18I6, genomic
            survey sequence.
ACCESSION   B95930
VERSION     B95930.1  GI:2997868
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 147)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
            Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T18I6TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 147.
FEATURES             Location/Qualifiers
     source          1. .147
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T18I6"
                     /sex="hermaphrodite"
                     /clone_lib="TAMU"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
ORIGIN

  Query Match            100.0%;  Score 12;  DB 9;  Length 147;
  Best Local Similarity  100.0%;  Pred. No. 2.5e+04;
```

```
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         14 AAGGTTATATAT 3


RESULT 17
CA364088/c
LOCUS       CA364088                148 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  638929 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT72C17_A_B09 5',
            mRNA sequence.
ACCESSION   CA364088
VERSION     CA364088.1  GI:24672957
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 148)
  AUTHORS   Rexroad,C.E. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
            Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
  TITLE     Sequence analysis of a rainbow trout cDNA library and creation of a
            gene index
  JOURNAL   Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
COMMENT     Contact: Rexroad CE
            USDA, ARS, National Center for Cool and Cold Water Aquaculture
            11876 Leetown Road, Kearneysville, WV 25430, USA
            Tel: 304 724 8340 x2129
            Fax: 304 725 0351
            Email: crexroad@ncccwa.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329.
            Seq primer: AGCGGATAACAATTTCACACAGGA.
FEATURES             Location/Qualifiers
     source          1. .148
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="1RT72C17_A_B09"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="NCCCWA 1RT"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from brain, gill, liver,
                     spleen, muscle, and kidney."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 148;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        131 AAGGTTATATAT 120


RESULT 18
BG929548
LOCUS       BG929548                150 bp    mRNA    linear   EST 06-JUN-2001
DEFINITION  950006A12.x1 950 - Mature pollen from Sheila McCormick's lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BG929548
VERSION     BG929548.1  GI:14324071
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 950006  row: A  column: 12.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /tissue_type="pollen"
                     /dev_stage="mature"
                     /lab_host="SOLR"
                     /clone_lib="950 - Mature pollen from Sheila McCormick's
                     lab"
                     /note="Vector: Stratagene's Uni-Zap XR (pBluescript SK-);
                     Site_1: EcoRI (5-prime); Site_2: XhoI (3-prime);
                     Unamplified cDNA library directionally cloned by Rima
                     Kulikauskas using Stratagene's Uni-Zap system. Insert
                     sizes ranged from 0.5Kb to 2Kb.  50 microliter aliquot had
                     338,000 pfu when it was made in Sept, 1995,  from oligo
                     dT-primed poly A+ RNA."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         59 AAGGTTATATAT 70


RESULT 19
AV107977
LOCUS       AV107977                153 bp    mRNA    linear   EST 28-JUN-1999
DEFINITION  AV107977 Mus musculus liver C57BL/6J 13-day embryo Mus musculus
            cDNA clone 2510042D07, mRNA sequence.
ACCESSION   AV107977
VERSION     AV107977.1  GI:5255525
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 153)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
```

trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998)) Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998)) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .153
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2510042D07"
/sex="mixed"
/tissue_type="liver"
/dev_stage="13-day embryo"
/clone_lib="Mus musculus liver C57BL/6J 13-day embryo"

ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 52 AAGGTTATATAT 63

RESULT 20
BE058406/c
LOCUS BE058406 155 bp mRNA linear EST 14-JUL-2004
DEFINITION sn15f09.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1016-11586 5', mRNA sequence.
ACCESSION BE058406
VERSION BE058406.1 GI:8402772
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 155)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Insert Length: 344 Std Error: 0.00
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .155
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Williams 82"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1016-11586"
/tissue_type="immature flowers of field grown plants"
/lab_host="XL10-Gold"
/clone_lib="Gm-c1016"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

ORIGIN

Query Match 100.0%; Score 12; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 41 AAGGTTATATAT 30

RESULT 21
CA766087
LOCUS CA766087 157 bp mRNA linear EST 08-JAN-2003
DEFINITION AF53-Rpf_10_A15_T7_050.ab1 IRRI Drought Stress Panicle Library Oryza sativa (indica cultivar-group) cDNA clone C0003471 5' similar to unknown, mRNA sequence.
ACCESSION CA766087
VERSION CA766087.2 GI:27548005
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 157)
AUTHORS Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and Bruskiewich,R.M.
TITLE IRRI Drought Stress Panicle cDNA Library
JOURNAL Unpublished (2002)
COMMENT On Dec 2, 2002 this sequence version replaced gi:25995342.
Contact: Richard Bruskiewich
Biometrics and Bioinformatics Unit
International Rice Research Institute
DAPO 7777, Metro Manila, Philippines
Tel: +63-2-845-0563
Fax: +63-2-845-0606
Email: r.bruskiewich@cgiar.org
International Rice Information System (IRIS; http://www.iris.irri.org): D0203470
Assignment of putative function to the sequence by S. Rudd of the Munich Information Center for Protein Sequences (http://mips.gsf.de)
Plate: 10 row: A column: 15.
FEATURES Location/Qualifiers
source 1. .157
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="C0003471"
/tissue_type="Panicles"
/dev_stage="Flowering"
/clone_lib="IRRI Drought Stress Panicle Library"
/note="Vector: pBluescript II SK+; Water stress was applied by not watering for 4 consecutive days. Panicles were collected from control (well watered) and stressed plants at 2 days before heading, at heading, 50% flowering

and 4 days after 50% flowering."
ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 157;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 101 AAGGTTATATAT 112

RESULT 22
CO720276/c
LOCUS CO720276 158 bp mRNA linear EST 27-JUL-2004
DEFINITION tai41f12.y1 HyEch JUMY T1 Hydractinia echinata cDNA 5', mRNA sequence.
ACCESSION CO720276
VERSION CO720276.1 GI:50697149
KEYWORDS EST.
SOURCE Hydractinia echinata
ORGANISM Hydractinia echinata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae; Hydractiniidae; Hydractinia.
REFERENCE 1 (bases 1 to 158)
AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WashU Hydra EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Hans Bode
WashU Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Marcus Frohme and Uri Frank Library materials provided by Marcus Frohme, German Cancer Research Center(DKFZ) Heidelberg, Uri Frank, University of Heidelberg Library re-arrayed by Jorge Sozaried (DKFZ) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hrbode@uci.edu)
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .158
/organism="Hydractinia echinata"
/mol_type="mRNA"
/db_xref="taxon:35630"
/lab_host="DH10B"
/clone_lib="HyEch JUMY T1"
/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; A pool of mRNA was primed with an anchored oligo-dT adaptor with a Not I site for 1st strand synthesis. Double stranded cDNA was ligated to Sal-Adaptors and cut with NotI. After size selection (without radioactivity) cDNA was ligated into pSport with SalI/NotI termini. Transformation was in DH10b Electromax T1 phage resistant cells. Plating was on 2YT/Carbenicillin Agar and 14x384 cells were blue/white selected automatically picked into 2YT/HMFM/Carbenicillin (HMFM is a freezing additive)."
ORIGIN

Query Match 100.0%; Score 12; DB 7; Length 158;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 81 AAGGTTATATAT 70

RESULT 23
T26907/c
LOCUS T26907 159 bp mRNA linear EST 14-SEP-1998
DEFINITION ESTDIR575 CD34+DIRECTIONAL Homo sapiens cDNA clone CDDIR575 3', mRNA sequence.
ACCESSION T26907
VERSION T26907.1 GI:567798
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159)
AUTHORS Yang,Y., Peterson,K.R., Stamatoyannopoulos,G. and Papayannopoulou,T.
TITLE Human CD34+ cell EST database: single-pass sequencing of 402 clones from a directional cDNA library
JOURNAL Exp. Hematol. 24 (5), 605-612 (1996)
PUBMED 8605965
COMMENT Contact: Yi Yang
Medical Genetics, RG-25
University of Washington
Health Science Building, Room BB512, Seattle, WA 98195
Tel: 2065432368
Fax: 2065433050
Email: yi629@u.washington.edu
Seq primer: T7 promoter.
FEATURES Location/Qualifiers
source 1. .159
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CDDIR575"
/clone_lib="CD34+DIRECTIONAL"
/note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human CD34+ cells. cDNA was synthesized using Stratagene's ZAP-cDNA synthesis kit(Cat#200400). cDNA was directionally cloned into EcoRI and XhoI sites in vector Lambda ZAPII"
ORIGIN

Query Match 100.0%; Score 12; DB 8; Length 159;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 29 AAGGTTATATAT 18

RESULT 24
CD162149
LOCUS CD162149 160 bp mRNA linear EST 14-SEP-2003
DEFINITION ML1-0075T-R173-D01-U.G ML1-0075 Schistosoma mansoni cDNA clone ML1-0075T-R173-D01.G, mRNA sequence.
ACCESSION CD162149
VERSION CD162149.1 GI:34698864
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 160)
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,

```
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0075T-R173  row: 1  column: D.
FEATURES             Location/Qualifiers
     source          1. .160
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0075T-R173-D01.G"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0075"
                     /note="Vector: pGEM T-easy"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 160;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        132 AAGGTTATATAT 143


RESULT 25
BG513953/c
LOCUS       BG513953                 161 bp    mRNA    linear   EST 28-MAR-2001
DEFINITION  dac15h11.y2 NICHD_XGC_He1 Xenopus laevis cDNA clone IMAGE:4407404
            5', mRNA sequence.
ACCESSION   BG513953
VERSION     BG513953.1  GI:13484610
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 161)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L.,
            Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y.,
            Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .161
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4407404"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_He1"
                     /note="Organ: heart; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 161;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        126 AAGGTTATATAT 115


RESULT 26
CE584959
LOCUS       CE584959                 163 bp    DNA     linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-17000366332200 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE584959
VERSION     CE584959.1  GI:36901740
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          7 AAGGTTATATAT 18


RESULT 27
BX655471/c
LOCUS       BX655471                 165 bp    DNA     linear   GSS 04-APR-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-589E01-021428,
```

```
            genomic survey sequence.
ACCESSION   BX655471
VERSION     BX655471.1  GI:37611859
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
  TITLE     GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for
            the identification of T-DNA insertion mutants in Arabidopsis
            thaliana
  JOURNAL   Bioinformatics 19 (11), 1441-1442 (2003)
   PUBMED   12874060
REFERENCE   2
  AUTHORS   Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and
            Weisshaar,B.
  TITLE     An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
            flanking sequence tag-based reverse genetics
  JOURNAL   Plant Mol. Biol. 53 (1-2), 247-259 (2003)
   PUBMED   14756321
REFERENCE   3
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and
            Weisshaar,B.
  TITLE     High-throughput generation of sequence indexes from T-DNA
            mutagenized Arabidopsis thaliana lines
  JOURNAL   BioTechniques 35 (6), 1164-1168 (2003)
   PUBMED   14682050
REFERENCE   4  (bases 1 to 165)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence has been recovered from the left border of the T-DNA.
            It indicates an insertion within the locus defined by BAC clone
            fca_all. Details on the protocols used for generation of the
            sequence are described in References 1-3. The sequences are
            generated at the MPI for Plant Breeding Research in the context of
            the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
            program designated 'GABI'. Information on line availability can be
            found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .165
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="GK-589E01-021428"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Col-0"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161 (GenBank accession number: AJ537514). The
                     lines contain one or more T-DNA insertions. The DNA
                     fragment(s) resulting from the PCR were directly sequenced
                     to determine the genomic sequence flanking the insertion.
                     T-DNA derived sequences were removed."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 165;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        129 AAGGTTATATAT 118


RESULT 28
CV314160/c
LOCUS       CV314160                166 bp    mRNA    linear   EST 24-SEP-2004
DEFINITION  CM1-FN0103-030700-296-b01 FN0103 Homo sapiens cDNA, mRNA sequence.
ACCESSION   CV314160
VERSION     CV314160.1  GI:52637374
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 166)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. http://www.ludwig.org.br.
FEATURES             Location/Qualifiers
     source          1. .166
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="FN0103"
                     /note="Organ: prostate_normal; Vector: puc18; Site_1:
                     SmaI; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 7;  Length 166;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        155 AAGGTTATATAT 144


RESULT 29
BF911105
LOCUS       BF911105                168 bp    mRNA    linear   EST 18-JAN-2001
DEFINITION  PM1-UT0061-061100-010-a06 UT0061 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF911105
VERSION     BF911105.1  GI:12302563
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 168)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
```

```
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-UT0061-
            061100-010-a06&t3=2000-11-06&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 135.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UT0061"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        132 AAGGTTATATAT 143


RESULT 30
BP616314/c
LOCUS       BP616314                 168 bp    mRNA    linear   EST 26-JUN-2004
DEFINITION  BP616314 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-19-I08 3',
            mRNA sequence.
ACCESSION   BP616314
VERSION     BP616314.1  GI:49267496
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 168)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-19-I08"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 3;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        114 AAGGTTATATAT 103


Search completed: December 20, 2005, 15:07:57
Job time : 68.7751 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:28:07 ; Search time 3.01606 Seconds
(without alignments)
5614.164 Million cell updates/sec

Title: US-09-837-344-42
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4324718

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22916/app_query.fasta_1.3740
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_1319_@runat_21122005_093215_22916 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 11:26:19
Job time : 3.01606 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 250.224 Seconds
(without alignments)
5204.614 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22895/app_query.fasta_1.3740
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_821_@runat_21122005_093215_22895 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2475	99.5	493	2	AAR26944	Aar26944 P.falcipa
2	1637.5	65.8	456	8	ADO21942	Ado21942 LSA-NRC(H
3	1627	65.4	457	8	ADO21920	Ado21920 LSA-NRC(H
4	1558	62.6	316	2	AAR26941	Aar26941 P.falcipa
5	972.5	39.1	462	2	AAR05766	Aar05766 Portion o
6	770.5	31.0	318	2	AAR26943	Aar26943 P.falcipa
7	600	24.1	117	2	AAR26937	Aar26937 P.falcipa
c 8	376.5	14.3	419	4	ABG17301	Abg17301 Novel hum
9	376	15.1	1898	2	AAY30795	Aay30795 A human t
10	376	15.1	1898	7	ADD48869	Add48869 Human Pro
11	374	15.0	1439	8	ADR12189	Adr12189 Trichohya
12	374	15.0	1439	8	ADU77050	Adu77050 Murine ha
13	374	15.0	1439	9	ADW00404	Adw00404 Epithelia
14	373	15.0	1584	8	ABM80133	Abm80133 Tumour-as
15	362.5	14.6	412	2	AAW03626	Aaw03626 Human thy
16	359	14.4	1162	3	AAY96255	Aay96255 Kaposi's
17	359	14.4	1162	3	AAY58500	Aay58500 HHV8 ORF
18	359	14.4	1162	4	AAB62331	Aab62331 Amino aci
19	359	14.4	1162	5	ABB05621	Abb05621 Kaposi's
20	359	14.4	1162	8	ADJ65096	Adj65096 HHV8 late
21	359	14.4	1162	9	ADV68154	Adv68154 Kaposi's
22	339.5	13.7	207	3	AAB44672	Aab44672 Gene 33 h
23	335	13.5	562	2	AAR70491	Aar70491 Leucocyto
24	334.5	13.4	1192	3	AAB18165	Aab18165 Plasmodiu
25	330.5	13.3	2990	8	ADP29884	Adp29884 Human sec
26	330	13.3	2274	4	ABB58657	Abb58657 Drosophil
27	328.5	13.2	360	2	AAW03627	Aaw03627 Human fol
28	325	13.1	345	4	ABU53267	Abu53267 Human tes
29	325	13.1	611	2	AAY29039	Aay29039 T. gondii
30	325	13.1	611	4	AAU25510	Aau25510 T. gondii
31	325	13.1	611	7	ADG17131	Adg17131 T. gondii
32	321	12.9	658	4	ABB65632	Abb65632 Drosophil
33	321	12.9	800	4	ABB71459	Abb71459 Drosophil
34	318	12.8	554	4	ABB59454	Abb59454 Drosophil
35	318	12.8	3616	8	ADF45530	Adf45530 Chicken A
36	317.5	12.8	1937	8	ADQ17289	Adq17289 Human sof
37	317.5	12.8	1937	8	ADQ17242	Adq17242 Human sof
38	317	12.7	1972	2	AAW00024	Aaw00024 Smooth mu
39	316	12.7	962	7	ADE78228	Ade78228 Endometri
40	316	12.7	1180	7	ADF06151	Adf06151 Bacterial
41	316	12.7	1976	6	ABU07447	Abu07447 Protein d
42	316	12.7	1976	9	ADY16758	Ady16758 PRO polyp
43	316	12.7	1988	4	AAM41000	Aam41000 Human pol
44	316	12.7	1988	4	AAM40999	Aam40999 Human pol
45	315.5	12.7	1170	5	ABB05710	Abb05710 Human tes

ALIGNMENTS

RESULT 1
AAR26944
ID AAR26944 standard; protein; 493 AA.
XX
AC AAR26944;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA gene C-terminal region.
XX
KW Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW paludism; liver stage-specific antigen.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT Region 13. .213
FT /label= repetitive_region
FT Region 214. .493
FT /label= non-repetitive_region
XX
PN WO9213884-A1.
XX
PD 20-AUG-1992.
XX
PF 05-FEB-1992; 92WO-FR000104.
XX
PR 05-FEB-1991; 91FR-00001286.
XX
PA (INSP) INST PASTEUR.
XX
PI Guerinmarchand C, Druilhe P;

```
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28119.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 2; Fig 8-10; 81pp; French.
XX
CC   The 3' part of the P.falciparum liver-stage specific antigen (LSA) gene
CC   codes for a polypeptide sequence which carries a T cell epitope
CC   characteristic of a protein produced in hepatocytes infected with
CC   P.falciparum. The polypeptide can be used in the preparation of vaccines
CC   against malaria. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 493 AA;

Alignment Scores:
Pred. No.:              8.03e-177       Length:         493
Score:                  2475.00         Matches:        492
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.80%          Mismatches:     0
Query Match:            99.52%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x AAR26944 (1-493)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 20

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 40

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 60

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 80

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGlu 100

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGln 120

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 140

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLys 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu 240

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 260

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 280

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 300

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 320

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGln 340

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGlu 360

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAsp 380

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeu 400

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLys 420

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeu 440

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 TyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLys 460

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 PheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAsp 480

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||
Db        481 GluLeuSerGluAspIleThrLysTyrPheMetLysLeu 493

RESULT 2
ADO21942
ID   ADO21942 standard; protein; 456 AA.
XX
AC   ADO21942;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H) construct protein derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H) construct.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
PN   WO2004044167-A2.
XX
```

```
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   N-PSDB; ADO21941.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 4; SEQ ID NO 26; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to a C-terminal His6 tag.
XX
SQ   Sequence 456 AA;

Alignment Scores:
Pred. No.:                3.71e-114      Length:          456
Score:                    1637.50        Matches:         337
Percent Similarity:       79.65%         Conservative:    27
Best Local Similarity:    73.74%         Mismatches:      70
Query Match:              65.84%         Indels:          23
DB:                       8              Gaps:            5

US-09-837-344-45 (1-1482) x ADO21942 (1-456)

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||::::::   :::   |||      |||:::        |||:::
Db          6 GluLysAspGluIleIleLysSerAsnLeuArgSerGlySerSerAsn------SerArg 23

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
                |||      |||:::      |||::::::      |||
Db         24 AsnArgIleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerTyrGlu 43

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---------------- 288
              |||   |||      |||             ::::::|||:::
Db         44 LysThrLysAsnAsnGluAsnAsnLysPhePheAspLysAspLysGluLeuThrMetSer 63

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                 |||         |||                 |||   :::
Db         64 AsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer 83

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
              |||   :::      :::::::::   |||   :::|||   :::      |||   :::
Db         84 GluAsnIlePheLeuLysGluAsnLysLeuAsnLysGluGlyLysLeuIleGluHisIle 103

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
                      |||      :::::::::      |||   :::   :::   :::
Db        104 IleAsnAspAspAsp------LysLysLysTyrIleLysGlyGlnAspGluAsnArg 121

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
              |||   |||||||||                |||||      |||||||||||||||
Db        122 GlnGluAspLeuGlu-----------------GluLysAlaAlaGluGlnGlnSerAsp 135

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
              ||||||||||||||   ||||||||||||||                          |||
```

```
Db        136 LeuGluGlnGluArgLeuAlaLysGluLysLeu-----------------------Gln 147

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAla 167

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        168 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAlaGluAspLeu 187

Qy        709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 207

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 IleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGlu 227

Qy        829 ATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg 247

Qy        889 GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAA 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 AspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLys 267

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 GluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspVal 287

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 AsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeu 307

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 IleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAsp 327

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLys 347

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGlu 367

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsn 387

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLys 407

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeu 427

Qy       1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 444

RESULT 3
ADO21920
ID   ADO21920 standard; protein; 457 AA.
XX
AC   ADO21920;
XX
DT   12-AUG-2004  (first entry)
XX
DE   LSA-NRC(H)Mut construct protein derived from Malaria parasite LSA-1.
XX
KW   LSA-NRC; liver stage antigen; LSA-1; T-cell; B-cell epitope;
```

```
KW   parasitophorous vacuole; antimalarial; vaccine; immunotherapy;
KW   malaria parasite P. falciparum; LSA-NRC(H)Mut construct; mutant; mutein.
XX
OS   Plasmodium falciparum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 433
FT                   /note= "Residue inserted relative to wild-type P.
FT                   falciparum T5 epitope"
XX
PN   WO2004044167-A2.
XX
PD   27-MAY-2004.
XX
PF   12-NOV-2003; 2003WO-US036011.
XX
PR   12-NOV-2002; 2002US-0425719P.
XX
PA   (REED-) REED ARMY INST RES WALTER.
XX
PI   Lanar DE,  Hillier CJ,  Lyon JA,  Angov E,  Kumar S,  Rogers W;
PI   Barbosa A;
XX
DR   WPI; 2004-420309/39.
DR   N-PSDB; ADO21919.
XX
PT   Recombinant LSA-NRC polypeptide for use as a vaccine and a diagnostic
PT   agent, preferably for diagnosing malaria, comprises a liver stage antigen
PT   (LSA-1) epitope.
XX
PS   Claim 5; SEQ ID NO 4; 90pp; English.
XX
CC   The invention relates to a novel recombinant LSA-NRC polypeptide
CC   comprising liver stage antigen (LSA-1) T-cell and B-cell epitopes. LSA-1
CC   protein is found within the parasitophorous vacuole (PV) of Plasmodium
CC   falciparum. The polypeptide of the invention demonstrates antimalarial
CC   activities and may be useful as a vaccine during immunotherapy and as a
CC   diagnostic agent, preferably for diagnosing malaria. The current sequence
CC   is that of the LSA-NRC(H)Mut construct protein of the invention. The
CC   construct is harmonised for expression in Escherichia coli and comprises
CC   the harmonised N-terminal, C-terminal and 2 tandem repeats of the Malaria
CC   parasite P. falciparum LSA-1 in addition to an amino acid insertion
CC   within the T5 P. falciparum epitope and a C-terminal His6 tag.
XX
SQ   Sequence 457 AA;

Alignment Scores:
Pred. No.:              2.26e-113      Length:         457
Score:                  1627.00        Matches:        337
Percent Similarity:     79.48%         Conservative:   27
Best Local Similarity:  73.58%         Mismatches:     70
Query Match:            65.42%         Indels:         24
DB:                     8              Gaps:           6

US-09-837-344-45 (1-1482) x ADO21920 (1-457)

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||::::::   :::   |||      |||:::         |||:::
Db          6 GluLysAspGluIleIleLysSerAsnLeuArgSerGlySerSerAsn------SerArg 23

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
                  |||      |||:::      |||::::::      |||
Db         24 AsnArgIleAsnGluGluLysHisGluLysLysHisValLeuSerHisAsnSerTyrGlu 43

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA-------------- 288
              |||   |||      |||                :::::|||:::
Db         44 LysThrLysAsnAsnGluAsnAsnLysPhePheAspLysAspLysGluLeuThrMetSer 63

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                  |||         |||               |||   :::
Db         64 AsnValLysAsnValSerGlnThrAsnPheLysSerLeuLeuArgAsnLeuGlyValSer 83

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
              |||   :::      :::::::::   |||   :::|||   :::      |||   :::
Db         84 GluAsnIlePheLeuLysGluAsnLysLeuAsnLysGluGlyLysLeuIleGluHisIle 103

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
                          |||      :::::::::      |||   :::   :::   :::
Db        104 IleAsnAspAspAspAsp------LysLysLysTyrIleLysGlyGlnAspGluAsnArg 121

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
              |||   |||||||||                  ||||||      ||||||||||||||||
Db        122 GlnGluAspLeuGlu-----------------GluLysAlaAlaGluGlnGlnSerAsp 135

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
              ||||||||||||||   ||||||||||||||                          |||
Db        136 LeuGluGlnGluArgLeuAlaLysGluLysLeu----------------------Gln 147

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAla 167

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        168 AspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAlaGluAspLeu 187

Qy        709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyr 207

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 IleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGlu 227

Qy        829 ATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 IleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArg 247

Qy        889 GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAA 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 AspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLys 267

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 GluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspVal 287

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 AsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeu 307

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 IleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAsp 327

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 AsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLys 347

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 AsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGlu 367

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsn 387

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAATGATAAGCAGGTTAATAAG 1368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLys 407

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeu 427

Qy       1429 CAGATCGTGGATGAG---TTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
```

```
                ||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        428 GlnIleValAspGluArgLeuSerGluAspIleThrLysTyrPheMetLysLeu 445

RESULT 4
AAR26941
ID   AAR26941 standard; protein; 316 AA.
XX
AC   AAR26941;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum LSA-R-NR protein.
XX
KW   Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism; liver stage-specific antigen.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .209
FT                   /label= repeat_region
FT                   /note= "contains 12 x 17mer repeats"
FT   Region          210. .316
FT                   /label= non-repeat_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28115.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Disclosure; Fig 1; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert (950 bases) encoded LSA-R-NR
CC   containing a 12-repeat region followed by a non-repeat region. Preferred
CC   antigenic polypeptides of the invention are derived from the amino acid
CC   sequence of LSA-R-NR. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 316 AA;

Alignment Scores:
Pred. No.:              3.15e-108      Length:         316
Score:                  1558.00        Matches:        315
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  99.68%         Mismatches:     0
Query Match:            62.65%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-45 (1-1482) x AAR26941 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 5
AAR05766
ID   AAR05766 standard; protein; 462 AA.
XX
AC   AAR05766;
XX
DT   25-MAR-2003  (revised)
DT   05-NOV-1990  (first entry)
XX
DE   Portion of peptide antigen to malarial sporozite.
XX
KW   Malaria; sporozite; vaccine; exoerythrocytic parasites; tetanus toxoid.
XX
```

OS Plasmodium falciparum.
XX
PN WO9006130-A.
XX
PD 14-JUN-1990.
XX
PF 30-NOV-1988; 88US-00278234.
XX
PR 30-NOV-1988; 88US-00278234.
PR 12-APR-1989; 89US-00337204.
XX
PA (BIOM-) BIOMEDICAL RES INST.
XX
PI Hollingdal MR;
XX
DR WPI; 1990-209624/27.
DR N-PSDB; AAQ05140.
XX
PT Novel malarial sporozoite antigenic protein - useful as vaccine against
PT sporozoite(s) and exo:erythrocytic parasites.
XX
PS Disclosure; Page ?; -pp; English.
XX
CC Antigen, preferably linked at the C-terminal to a carrier such as tetanus
CC toxoid, may be used as a vaccine against the malarial sporozite. (Updated
CC on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 462 AA;

Alignment Scores:

Pred. No.:	2.14e-64	Length:	462
Score:	972.50	Matches:	232
Percent Similarity:	61.88%	Conservative:	57
Best Local Similarity:	49.68%	Mismatches:	95
Query Match:	39.10%	Indels:	83
DB:	2	Gaps:	10

US-09-837-344-45 (1-1482) x AAR05766 (1-462)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||   |||||||||||||||||||||||||   |||||||||||||||||||||||
Db         74 GlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 93

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db         94 GlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 113

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        114 LeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 133

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||||||||||||||||||||||||||||||||||||||||||||:::   :::
Db        134 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluArgThrLysAla 153

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              :::   |||   |||   |||||||||||||||||||||||:::|||||||||||||||
Db        154 SerThrGluThrLeuHisGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGlu 173

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              |||||||||   |||||||||||||||||||||||||   |||||||||||||||||||
Db        174 LysLeuGlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 193

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||   |||||||||||||||||||||||||||||
Db        194 GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGln 213

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        214 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 233

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||:::|||   |||||||||||||||||||||||||||||||||||||||||:::

Db        234 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAsp 253

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||||||||||||||||||||||||||||||||||||||||||||:::|||   |||
Db        254 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAla 273

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||||||||||||||   ||||||   |||||||||:::   ||||||   |||:::
Db        274 LysGluLysLeuGlnValGlnGlnSerAspLeuGluArgThrLysAlaSerThrGluThr 293

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                  ||||||
Db        294 LeuGlnGluArg---------------------------------------------- 297

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                                                                    |||
Db        298 ----------------------------------------------------------Gln 298

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||   |||   |||:::   |||         :::     ::::::   |||   :::
Db        299 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 318

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::     |||                    |||   :::|||||:::   :::
Db        319 GluGlnGluArgArgAlaLysGluLysLeuGln---GluGlnGlnArgAspLeuGluGln 337

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                  |||   :::                  |||:::   :::::::::
Db        338 GluArgLeuAlaLys----------------------GluLysLeuGlnGluGlnGln 349

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              :::|||::::::         |||         :::
Db        350 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeu---------------------- 362

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                                                                    :::
Db        363 ----------------------------------------------------------Gln 363

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||         ||||||::::::   :::      :::      :::      |||:::
Db        364 GluGlnGlnArgAspLeuGluGlnGluArg------LeuAlaLysGluLysLeuGlnGlu 381

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::::::::               :::||||||                |||   |||
Db        382 GlnGlnSer--------------AspLeuGluGlnGluArgArgAlaLysGluLysLeu 396

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
                 :::           :::|||:::      |||   |||::::::   |||::::::
Db        397 GlnGluGlnGlnSerAspLeuGluGln------GluArgLeuAlaLysValLysLeuGln 414

Qy       1261 AAAGGAAAGAAATAT---GAAAAAACAAAG------GATAATAATTTTAAACCAAATGAT 1311
              :::   :::   |||   |||:::||||||            :::
Db        415 GluGlnGlnSerTyrLeuGluArgThrLysAlaSerThrGluThrLeuGlnGluArgGln 434

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAA---AATGATAAGCAGGTTAATAAG 1368
                    |||      |||   :::   |||   |||      ::::::|||   :::
Db        435 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 454

Qy       1369 GAAAAGGAAAAATTCATAAAA 1389
              |||:::|||:::      |||
Db        455 GluGlnGluArgArgAlaLys 461
```

RESULT 6
AAR26943
ID AAR26943 standard; protein; 318 AA.
XX
AC AAR26943;
XX
DT 25-MAR-2003 (revised)
DT 08-FEB-1993 (first entry)
XX
DE P.falciparum LSA N-terminal portion.
XX

```
KW   Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW   paludism; liver stage-specific antigen; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .153
FT                   /label= unique_5'_region
FT   Region          154. .318
FT                   /label= repetitive_region
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
DR   N-PSDB; AAQ28117.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 11; Fig 7; 81pp; French.
XX
CC   The 5' part of the P.falciparum liver-stage specific antigen (LSA) gene
CC   codes for a polypeptide sequence which carries a T cell epitope
CC   characteristic of a protein produced in hepatocytes infected with
CC   P.falciparum. The polypeptide can be used in the preparation of vaccines
CC   against malaria. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 318 AA;

Alignment Scores:
Pred. No.:              2.67e-49        Length:         318
Score:                  770.50          Matches:        163
Percent Similarity:     92.59%          Conservative:   12
Best Local Similarity:  86.24%          Mismatches:     13
Query Match:            30.98%          Indels:         1
DB:                     2               Gaps:           1

US-09-837-344-45 (1-1482) x AAR26943 (1-318)

Qy          16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 75
               |||   ::::::::::::      |||   :::   :::   :::|||   |||||||||
Db         131 AspAspAspLysLysLysTyrIleLysGlyGlnAspGluAsnArgGlnGluAspLeuGlu 150

Qy          76 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 135
                 ::::::   ||||||||||||||||||   ||||||||||||   |||||||||||||
Db         151 ---GluLysAlaAlaLysGluLysLeuGlnGlyGlnGlnSerAspSerGluGlnGluArg 169

Qy         136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA 195
               ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         170 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLys 189

Qy         196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         190 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 209

Qy         256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
               ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db         210 GlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 229

Qy         316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         230 GlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAsp 249

Qy         376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         250 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 269

Qy         436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         270 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeu 289

Qy         496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         290 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 309

Qy         556 AAGTTGCAAGAACAACAAAGCGATTTA 582
               :::|||||||||||||||||||||||||
Db         310 ArgLeuGlnGluGlnGlnSerAspLeu 318

RESULT 7
AAR26937
ID   AAR26937 standard; peptide; 117 AA.
XX
AC   AAR26937;
XX
DT   25-MAR-2003  (revised)
DT   08-FEB-1993  (first entry)
XX
DE   P.falciparum liver-stage specific antigen non-repeat region.
XX
KW   Malaria; LSA-NR; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW   paludism.
XX
OS   Plasmodium falciparum.
XX
PN   WO9213884-A1.
XX
PD   20-AUG-1992.
XX
PF   05-FEB-1992;   92WO-FR000104.
XX
PR   05-FEB-1991;   91FR-00001286.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Guerinmarchand C,  Druilhe P;
XX
DR   WPI; 1992-299985/36.
XX
PT   Polypeptide(s) derived from liver stage of PLASMODIUM FALCIPARUM - for
PT   vaccination against, treatment of and diagnosis of malaria.
XX
PS   Claim 4; Page 55; 81pp; French.
XX
CC   A genomic DNA bank of P.falciparum EcoRI fragments prepared in lambda
CC   gt11 was used to transform E.coli. The expression library was screened
CC   with human antisera against antigens of all stages of P. falciparum. The
CC   library was rescreened with antibodies affinity- purified on a clone
CC   which was able to recognise antibodies specific to the hepatic phase.
CC   About 40 clones were detected which produced a characteristic LSA
CC   epitope. The clone with the largest insert encoded LSA-R-NR containing a
CC   12-repeat region (LSA-R) followed by a non-repeat region (LSA-NR). The
CC   sequence given here is a preferred polypeptide of the invention carrying
CC   a T cell epitope typical of liver-stage P.falciparum. See AAR26919-
CC   R26935, AAR26937 and AAR26939. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 117 AA;

Alignment Scores:
Pred. No.:              1.36e-36        Length:         117
Score:                  600.00          Matches:        116
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.15%          Mismatches:     0
Query Match:            24.13%          Indels:         0
DB:                     2               Gaps:           0
```

```
US-09-837-344-45 (1-1482) x AAR26937 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 8
ABG17301
ID   ABG17301 standard; protein; 419 AA.
XX
AC   ABG17301;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #17292.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS81488.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 47660; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 419 AA;

Alignment Scores:
Pred. No.:              8.36e-20        Length:         419
Score:                  376.50          Matches:        149
Percent Similarity:     44.34%          Conservative:   35
Best Local Similarity:  35.90%          Mismatches:     76
Query Match:            14.25%          Indels:         155
DB:                     4               Gaps:           25

US-09-837-344-45 (1-1482) x ABG17301 (1-419)

Qy       1217 CCTTCATCTAAATCATCTAAATTTTCATTTTTCTCTATCAAATCTTCTAGTTCTTTATAA 1158
              |||:::     ||||||     |||     |||:::   ||||||   ||||||
Db        136 ProAlaProAlaSerSerProCysSer---ArgSerLeuSerSerSerArgSerLeu--- 153

Qy       1157 ATTCCTATGTTTTCTTCATCTTGGAAA-----------------TTGTCATATTGCACA 1116
                          |||||||||   :::                 ||||||   |||
Db        154 -----------SerSerSerCysArgValSerCysCysHisSerLeuSerSerCysCys 169

Qy       1115 ATAGGCTTAAATTCGTCTAAGTCTTCATCATCTTCTTCTTCATCTATTAATGAATCGTCA 1056
                |||:::   |||:::|||||||||:::||||||   |||   :::   ||||||
Db        170 ArgSerLeuSerCysSerArgSerSerSerThrSerSerArgSerSerSerThrSerSer 189

Qy       1055 TATTCAGCACTTATTTCATCCTCATACTTACTTATTTGAAAATCATTAACATCAGAAATA 996
                |||      :::|||||||||
Db        190 SerSerLeuSerLeuSerSerSerLeu---------------------------------- 198

Qy        995 TTTACTGTCTCTAATGTATGATTTTGTATGTCAGCAGATTTATCTTCTTTTTGTTCTGGT 936
                                        |||   |||      ||||||      ||||||
Db        199 --------------------SerCysSerSerCysArgLeuSer------CysSer--- 208

Qy        935 TTTATTGAACCATCTTTCTTTTCTTCAAGATGTCCTTTATGTATATCCCTTCGTCCTTCA 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 ---------------------------------------------------------- 208

Qy        875 ACATTTGTTGTAATAGATTCTCTATTTGTTTTTTCTATTATAGACATTTCCTTGGAATCT 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 ---------------------------------------------------------- 208

Qy        815 CTACTATTCCCTCTGTTGTCCTGAGGTAAAGAAGATTGATGTGGTATATAATATCCACGT 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 ----------------------------------------------------------Asn 209

Qy        755 TCATTTTCTGATGGAAGTTCTATAGCTGGTATTTCTAAACGACCATATAAATCCTCTGCT 696
              |||||||||      ||||||         :::|||
Db        210 SerPheSerThrSerSerSerSerLeuSerLeuSer----------------------- 221

Qy        695 AATATATCTCCATGTTCCTTTTTTCTTTCTAAATTTTTTTTCGTATCAGCCTTCCTTTGT 636
                                                          |||:::      |||
Db        222 ----------------------------------------------SerSerLeuSerCys 226

Qy        635 TCTAAATCTCTTTGCTGCTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCG 576
              |||   |||   |||   |||||||:::   |||   :::      |||   |||   |||
Db        227 Ser---SerCysCysLeuSerCysSerAsnSerPheSerThrSerSerSerSerLeuSer 245

Qy        575 CTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGT 516
              |||      |||   :::   |||      ||||||||||||||   |||
Db        246 LeuSerSerSerLeuSerCysSerSerCysArgLeuSerCysSerAsnSerPhe------ 263
```

```
Qy        515 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAAC 456
                  :::   |||   :::   |||||||||      |||      ||||||   :::
Db        264 ------SerThrSerSerSerSerLeuSerCysProGlySerPheSerCysSerHisSer 281

Qy        455 TTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTA 396
                |||            |||||||||:::||||||      ||||||:::   ||||||
Db        282 LeuSer-----------SerCysSerArgSerLeu---SerSerCysSerHisSerLeu 296

Qy        395 GCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTC 336
                :::      ||||||||||:::||||||      ||||||:::   ||||||:::
Db        297 Ser------SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------ 311

Qy        335 TCTTGTTCTAAATCGCTTTGCTGCTCTTGTAACTTTTCTTTAGCAAGTCTATCTTGTTCT 276
                |||||||||:::||||||      ||||||:::   ||||||:::      |||||||||
Db        312 SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSer 328

Qy        275 AAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTT 216
                :::||||||      ||||||:::   ||||||:::      |||||||||:::||||||
Db        329 ArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSerArgSerLeu 345

Qy        215 ---TGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTTTGTTGT 159
                   |||||||||:::   |||              ||||||||||||||||||
Db        346 SerSerCysSerCysSerLeuSer-----------SerCysSerArgSerLeu---Ser 360

Qy        158 TCTTGCAACTTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGCTGCTCTTGTAAC 99
                |||||:::   |||||||:::      |||||||||   ||||||      ||||||:::
Db        361 SerCysSerHisSerLeuSer------SerCysSerHisSerLeu---SerSerCysSer 377

Qy         98 TTTTCTTTAGCAAGTCTATCTTGTTCTAAATCGCTTTGT-----------------TGT 57
                ||||||:::|||   |||||||||   |||   |||                   |||
Db        378 ArgSerLeuSerSerCysSerCysSerLeuSerSerCysSerArgCysLeuSerSerCys 397

Qy         56 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTT 12
                ||||||:::   |||            ||||||||||||||||
Db        398 SerCysSerLeuSer-----------SerCysSerArgSerLeu 408
```

RESULT 9
AAY30795
ID AAY30795 standard; protein; 1898 AA.
XX
AC AAY30795;
XX
DT 25-NOV-1999 (first entry)
XX
DE A human trichohyalin (TRHY) protein.
XX
KW Human; trichohyalin; TRHY; protein; tissue structure; wound healing;
KW terminally differentiating epidermal tissue; proteinaceous gel;
KW breast implant.
XX
OS Homo sapiens.
XX
PN US5958752-A.
XX
PD 28-SEP-1999.
XX
PF 14-FEB-1997; 97US-00800644.
XX
PR 30-APR-1993; 93US-00056200.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Kim I, Chung S, Park S, Steinert PM, Lee S;
XX
DR WPI; 1999-561041/47.
DR N-PSDB; AAZ22301.
XX
PT Human trichohyalin useful for forming a proteinaceous gel that promotes
PT wound healing.
XX
PS Disclosure; Fig 3A-W; 126pp; English.
XX

CC The present sequence represents a human trichohyalin (TRHY) protein. The
CC protein is found in terminally differentiating epidermal tissue, and is
CC involved in forming the structural architecture of such tissue. The
CC trichohyalin protein is useful for forming a proteinaceous gel which may
CC then be used for healing wounds, or in breast implants
XX
SQ Sequence 1898 AA;

Alignment Scores:

Pred. No.:	1.09e-19	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	2	Gaps:	24

US-09-837-344-45 (1-1482) x AAY30795 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
                :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
                :::         |||::::::|||||   :::          |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                ||||||   |||||||||   :::          :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC------------------ 219
                ||||||||||||   :::          :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                   |||:::|||||||||   :::          :::|||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                |||   ||||||   :::|||:::::::::                  ||||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg------------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG------------------CAAGAACAACAAAGCGATTTA--- 378
                |||       |||:::|||                  :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                         |||||||||   |||       |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGluGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA---------------CAAAGC 474
                   ::::::||||||   :::|||:::   :::|||:::                :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
                :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluAlaArg 713

Qy        526 ----------------------------------GATTTAGAACAAGAG--------- 540
                                                 ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
                |||::::::|||       |||:::      |||:::                   :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
                ||||||::::::   |||||||||:::      |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793
```

```
Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
              |||   |||   |||             :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||   |||             :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::             |||      |||                |||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                        :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                  :::|||   :::   |||    :::      |||:::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                  :::   :::   :::|||::::::           :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                      :::              ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                 |||          |||:::   |||||||||::::::       :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                              :::      :::   :::|||            |||:::
Db        935 ------------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::      |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::           :::      :::|||             |||      ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022
```

RESULT 10
ADD48869
ID ADD48869 standard; protein; 1898 AA.
XX
AC ADD48869;
XX
DT 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX
DE Human Protein AAA65582, SEQ ID NO 14580.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
OS Unidentified.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; AAA65582.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Example 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (described in Table 3
CC of the specification) which is differentially expressed during pain.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1898 AA;

Alignment Scores:

Pred. No.:	1.09e-19	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	7	Gaps:	24

US-09-837-344-45 (1-1482) x ADD48869 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||         |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::||||||   :::             |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                  ||||||   |||||||||   :::         :::|||::::::       ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578
```

```
Qy      178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
            ||||||||||||   :::         :::||||||:::
Db      579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy      220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
            |||:::|||||||||   :::         :::||||||:::      ||||||
Db      599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy      280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
            |||   ||||||   :::|||:::::::::               ||||||||||||
Db      619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy      340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
            |||      |||:::|||                 :::|||::::::   :::|||
Db      634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy      379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                    ||||||||   |||      |||:::|||:::   :::|||   :::
Db      654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy      430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
               ::::::||||||   :::|||:::   :::|||:::               :::
Db      674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy      475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
            :::   |||:::|||||||||   :::|||      :::|||:::|||
Db      694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGluGlnAlaArg 713

Qy      526 -----------------------------------GATTTAGAACAAGAG--------- 540
                                                ||||||   |||
Db      714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy      541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
            |||::::::|||      |||:::      |||:::                     :::
Db      734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy      580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
               |||::::::||||||   :::   :::||||||
Db      754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy      616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
               ||||||::::::   |||||||||:::      |||   :::          |||
Db      774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy      673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
            |||   |||   |||             :::   |||       |||
Db      794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy      718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                    ::::::   |||   |||                 :::|||      :::
Db      811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy      769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
            :::|||   :::             |||      |||                  ||||||
Db      829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy      829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
            :::      :::|||:::      :::                     :::   :::
Db      849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy      868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                ::|||   :::   |||      :::         |||:::::::::|||
Db      869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy      925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                ::   :::   :::|||::::::           :::   |||||||||
Db      889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy      985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                    :::               ::::::::::::
Db      907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy     1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
```

```
                ||||            |||:::   |||||||||::::::         :::|||
Db      917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy     1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                    :::      :::   :::|||          |||:::
Db      935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy     1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
            |||   |||   |||   ::::::   |||   |||      |||   |||      |||
Db      949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy     1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
            |||          :::      |||   :::   :::|||   ||||||   ::::::
Db      967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy     1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
            ::::::          :::      :::|||                |||      ||||||
Db      987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy     1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
            |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db     1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022
```

RESULT 11
ADR12189
ID ADR12189 standard; protein; 1439 AA.
XX
AC ADR12189;
XX
DT 21-OCT-2004 (first entry)
XX
DE Trichohyalin related protein, SEQ ID 1.
XX
KW Hair growth; murine; trichohyalin.
XX
OS Mus musculus.
XX
PN WO2004063750-A1.
XX
PD 29-JUL-2004.
XX
PF 07-JAN-2004; 2004WO-JP000030.
XX
PR 08-JAN-2003; 2003JP-00001891.
XX
PA (SUME) SUMITOMO ELECTRIC IND LTD.
XX
PI Takebe K, Hirai Y;
XX
DR WPI; 2004-571498/55.
DR N-PSDB; ADR12190.
XX
PT Evaluating hair-promoting activity, by using antibody that recognizes
PT proteins (antigen) of epithelial newborn follicles.
XX
PS Claim 1; SEQ ID NO 1; 39pp; Japanese.
XX
CC The present invention relates to a method (M1) for evaluating hair growth
CC -promoting activity, by using an antibody that recognizes an antigen
CC (ADR12189) of epithelial newborn follicles. Also claimed is a kit for
CC evaluating hair growth-promoting activity, comprising the antibody of
CC (M1). The present sequence is a murine trichohyalin-related protein,
CC which was used as the antigen in the method of the invention.
XX
SQ Sequence 1439 AA;

Alignment Scores:

Pred. No.:	1.49e-19	Length:	1439
Score:	374.00	Matches:	143
Percent Similarity:	48.15%	Conservative:	130
Best Local Similarity:	25.22%	Mismatches:	170
Query Match:	15.04%	Indels:	124
DB:	8	Gaps:	25

US-09-837-344-45 (1-1482) x ADR12189 (1-1439)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
              :::|||::::::   :::|||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||      |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                 ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                    ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                 ||||||||||||      :::|||:::          |||||||||:::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||          |||   :::   |||||:::::::        |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              ::::::|||   :::::::::   ::::::||:::          ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                    :::|||:::|||:::::::::   |||      ||||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                    |||   :::|||:::|||:::::::::            ||||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::|||:::   :::||||||   ||||||          |||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   |||||||||:::::|||             |||   :::      :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                 :::          ||||||      :::   :::
Db        871 GluGln------------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
              :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
                 :::|||:::   :::   ::::::         |||                 :::
```

```
Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                        |||   |||||||||       |||   |||   ||||||::::::
Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
                        :::   |||:::   |||:::|||          :::|||
Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ------------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                                       |||   |||           :::   |||
Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
                 :::|||   :::   |||   |||:::       |||   ::::::
Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ------------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
                                                 |||   :::|||:::
Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                 ::::::||||||   :::|||:::          :::   |||          :::|||
Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
                 |||   :::          |||   :::|||   |||         |||
Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 :::   |||                 ::::::   :::   :::       |||
Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
                 |||:::::::::::::::   ::::::   |||:::   :::|||         :::
Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                 :::            :::   |||   :::|||   :::::::::   :::   |||:::
Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377
                 :::      :::|||:::|||
Db       1141 LysPheArgGluGluGlnGlu 1147
```

RESULT 12
ADU77050
ID ADU77050 standard; protein; 1439 AA.
XX
AC ADU77050;
XX
DT 10-FEB-2005 (first entry)
XX
DE Murine hair growth stimulator related protein.
XX
KW Hair growth stimulation; endocrine-gen.; cardiovascular-gen.;
KW antidiabetic; hepatotropic; antiinflammatory; gastrointestinal-gen.;
KW muscular-gen.; stimulator of hair growth; cardiovascular disease;
KW diabetes; liver disease; periodontal disease; gastrointestinal disease;
KW muscular dystrophy; cholecystitis.
XX
OS Mus musculus.
XX
PN WO2004101610-A1.
XX
PD 25-NOV-2004.
XX
PF 14-MAY-2004; 2004WO-JP006909.
XX
PR 16-MAY-2003; 2003JP-00138567.
PR 17-NOV-2003; 2003JP-00386185.
XX

PA (SUME) SUMITOMO ELECTRIC IND LTD.
XX
PI Hirai Y, Nakajima K;
XX
DR WPI; 2004-833974/82.
XX
PT Novel oligopeptide or its modified form useful for enhancing hair growth
PT or for treating cardiovascular diseases, diabetes, liver diseases,
PT periodontal disease, intestinal disease or muscular dystrophy.
XX
PS Disclosure; SEQ ID NO 11; 80pp; Japanese.
XX
CC The invention relates to an oligopeptide or its modified form, used for
CC promoting hair growth, comprising 5-100 amino acid residues. The
CC invention also relates to a composition comprising the oligopeptide, its
CC modified form or its salt, and a hair growth stimulant comprising the
CC oligopeptide, its modified form or its salt. The oligopeptide is useful
CC for enhancing hair growth or for treating cardiovascular diseases,
CC diabetes, liver diseases, periodontal disease, intestinal disease,
CC muscular dystrophy or cholecystitis. This sequence represents a murine
CC hair growth stimulation related protein of the invention.
XX
SQ Sequence 1439 AA;

Alignment Scores:

Pred. No.:	1.49e-19	Length:	1439
Score:	374.00	Matches:	143
Percent Similarity:	48.15%	Conservative:	130
Best Local Similarity:	25.22%	Mismatches:	170
Query Match:	15.04%	Indels:	124
DB:	8	Gaps:	25

US-09-837-344-45 (1-1482) x ADU77050 (1-1439)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
              :::|||::::::   :::|||||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||      |||   |||:::         |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                    ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                  ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                |||||||||||      :::|||:::          |||||||||:::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||          |||   :::   ||||||::::::          |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              :::::|||   :::::::::   ::::::|||:::          ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                    :::|||:::|||:::::::::   |||      ||||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                    |||   :::|||:::|||:::::::::          ||||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                  :::|||:::   :::||||||   ||||||          ||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
                :::   |||||||||:::::|||          |||   :::      :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                  :::            ||||||      :::   :::
Db        871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
                :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA---------------TCTATT 864
                  :::|||:::   :::   ::::::          |||                  :::
Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                        |||   |||||||||      |||   |||   ||||||::::::
Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
                        :::   |||:::   |||:::|||          :::|||
Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ----------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                                       |||   |||             :::   |||
Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
                :::|||   :::   |||   |||:::      |||   ::::::
Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
                                                   |||   :::|||:::
Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                :::::::||||||   :::|||:::          :::   |||          :::|||
Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
                |||   :::          |||   :::|||   |||         |||
Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                :::   |||                ::::::   :::   :::         |||
Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
                |||::::::::::::::   ::::::   |||:::   :::|||         :::
Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                :::            :::   |||   :::|||   :::::::::   :::   |||:::
Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140
```

```
Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377
              :::      :::|||:::|||
Db       1141 LysPheArgGluGluGlnGlu 1147

RESULT 13
ADW00404
ID   ADW00404 standard; protein; 1439 AA.
XX
AC   ADW00404;
XX
DT   07-APR-2005  (first entry)
XX
DE   Epithelial hair follicle related antigen, SEQ ID 157.
XX
KW   Endocrine-Gen.; Cardiovascular-Gen.; Vulnerary; Osteopathic; Cardiant;
KW   Gene Therapy; antibody production; antibody therapy; antigen;
KW   hair growth; arteriosclerosis; osteoporosis; bone injury; angina; burns;
KW   wound healing.
XX
OS   Mus musculus.
XX
PN   US2005008602-A1.
XX
PD   13-JAN-2005.
XX
PF   07-JAN-2004; 2004US-00754079.
XX
PR   05-JUN-2000; 2000JP-00166903.
PR   04-JUN-2001; 2001WO-JP004691.
PR   13-NOV-2001; 2001JP-00347338.
PR   13-NOV-2001; 2001JP-00347340.
PR   05-DEC-2001; 2001JP-00371175.
PR   05-DEC-2001; 2001JP-00371366.
PR   23-MAY-2002; 2002US-00155922.
PR   08-JAN-2003; 2003JP-00001891.
XX
PA   (SUME ) SUMITOMO ELECTRIC IND LTD.
XX
PI   Hirai Y,  Takebe K;
XX
DR   WPI; 2005-080487/09.
DR   N-PSDB; ADW00406.
XX
PT   New antibody specifically recognizing an antigen of 220kDa present in
PT   epithelial new follicles, useful for evaluating and inducing hair growth
PT   promoting activity and morphogenesis.
XX
PS   Claim 5; SEQ ID NO 157; 87pp; English.
XX
CC   The present invention relates to a novel antibody, or its fragment, which
CC   specifically recognizes an antigen of about 220kDa (ADW00404) present in
CC   new epithelial hair follicles. The antigen is specifically expressed
CC   during the growth period of an imago or the developing period of a fetus.
CC   Also claimed is a method for the evaluation of hair growth promoting
CC   activity, comprising incubating skin tissue derived from a mammal in the
CC   presence of a substance to be tested to promote hair growth, recovering
CC   the skin tissue, reacting the skin tissue with the antibody cited above,
CC   and detecting the antibody, or its fragment, that reacts with the skin
CC   tissue. The method and antibody of the invention are useful for
CC   evaluating and inducing hair growth promoting activity, treating and
CC   ameliorating symptoms of diseases associated with abnormal morphogenesis,
CC   inducing cardiovascular revascularization, regeneration and endothelial
CC   cell growth, such as in arteriosclerosis, Berger's disease, osteoporosis,
CC   fractures and angina pectoris, and in treating burns or wounds.
XX
SQ   Sequence 1439 AA;

Alignment Scores:
Pred. No.:              1.49e-19      Length:          1439
Score:                  374.00        Matches:         143
Percent Similarity:     48.15%        Conservative:    130
Best Local Similarity:  25.22%        Mismatches:      170
Query Match:            15.04%        Indels:          124

DB:                  9             Gaps:          25

US-09-837-344-45 (1-1482) x ADW00404 (1-1439)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------------ 48
              :::|||::::::   :::||||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                |||:::|||:::|||        |||      |||   |||:::        |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA------------------------------------------ 114
                  ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                  ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
              ||||||||||||      :::|||:::         |||||||||::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||         |||   :::   ||||||::::::         |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              ::::::|||   :::::::::   ::::::|||:::         ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                   |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755

Qy        391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
                   |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy        442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
                   :::|||:::|||:::::::::   |||      ||||||||||||
Db        776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy        493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
                   |||   :::|||:::|||:::::::::         ||||||||||||
Db        796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy        544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                   |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::|||:::   :::||||||   ||||||         |||:::   :::   :::
Db        836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::   |||||||||:::::|||            |||   :::      :::   |||
Db        854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                 :::            ||||||      :::   :::
Db        871 GluGln-----------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT----------------- 819
              :::      |||   ::::::   ||||||      ||||||
Db        887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy        819 ------------------------------------------------------------ 819

Db        907 GlnLeuArgGlyGlnArgGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
```

```
                ::::|||:::   :::   ::::::         |||                    :::
Db        927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                  |||   |||||||||       |||   |||   ||||||::::::
Db        947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
                  :::   |||:::   |||:::|||         :::|||
Db        965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy        979 ------------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                                       |||   |||           :::   |||
Db        982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
                :::|||   :::   |||   |||:::       |||   ::::::
Db       1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy       1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
                                                |||   :::|||:::
Db       1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy       1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                :::::|||||   :::|||:::         :::   |||         :::|||
Db       1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy       1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
              |||   :::             |||   :::|||   |||         |||
Db       1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              :::   |||             ::::::   :::   :::         |||
Db       1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              |||:::::::::::::::   ::::::   |||:::   :::|||         :::
Db       1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy       1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::           :::   |||   :::|||   :::::::::   :::   |||:::
Db       1121 AspGlnGluLeuArgGlnGlyLeuGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy       1357 CAGGTTAATAAGGAAAAGGAA 1377
              :::      :::|||:::|||
Db       1141 LysPheArgGluGluGlnGlu 1147
```

RESULT 14
ABM80133
ID ABM80133 standard; protein; 1584 AA.
XX
AC ABM80133;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO80641, SEQ:339.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR N-PSDB; ACN37449.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 12; SEQ ID NO 339; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 1584 AA;

Alignment Scores:

Pred. No.:	1.79e-19	Length:	1584
Score:	373.00	Matches:	133
Percent Similarity:	46.53%	Conservative:	115
Best Local Similarity:	24.95%	Mismatches:	185
Query Match:	15.00%	Indels:	100
DB:	8	Gaps:	21

US-09-837-344-45 (1-1482) x ABM80133 (1-1584)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||         |||||||||   |||   :::::::::|||:::   |||
Db        278 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 297

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::|||||   :::             |||||||||
Db        298 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 314

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG---CAAGAACAACAAAGCGATCTAGAA 177
                 ||||||||         ::::::   |||   :::|||::::::      ||||||
Db        315 ArgGluGlnGluGluArgArgGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 334

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
              ||||||||||||   :::         :::||||||:::
Db        335 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 354

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                  |||:::|||||||||   :::         :::||||||:::      ||||||
Db        355 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 374

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGCGATTTAGAACAA 333
              |||   ||||||   :::|||:::::::::      ||||||:::         ||||||
Db        375 Gln---ArgLeuLysArgGluGluValGluArgLeuGluGlnGluGluArgArgGluGln 393
```

```
Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA-----------GAA 381
Db        394 ArgLeuLysArgGluGluProGluGluGluArgArgGlnGlnLeuLeuLysSerGluGlu 413

Qy        382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
Db        414 GlnGluGluArgArgGlnGlnGlnLeuArgArgGluGlnGlnGluArgArgGluGlnArg 433

Qy        442 CTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGCGATTTAGAACAA 486
Db        434 LeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArgGluHisGluGlu 453

Qy        487 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC-------------------- 525
Db        454 GluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArgGluArgIleLys 473

Qy        526 ----------------------GATTTAGAACAAGAG-------------------- 540
Db        474 SerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnSerLys 493

Qy        541 ---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591
Db        494 ValTyrSerArgProArgLysGlnGluGlyGlnArgArgArgGlnGluGlnGluGluLys 513

Qy        592 AGACGTGCTAAA-------------------------GAAAAGTTGCAAGAGCAGCAA 624
Db        514 ArgArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHisArgGlnGln 533

Qy        625 AGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGAAAAAAGGAA 681
Db        534 GlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGluLysSerGlu 553

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGT----------------------- 717
Db        554 ArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGluGlnArgGlu 570

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db        571 ArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------PheLeuProGlu 588

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
Db        589 GluGluGluLysGluGlnArgArgArgGlnArgArgGluArgGluLysGluLeuGlnPhe 608

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        609 LeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGlnGluGluGlu 628

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db        629 AspGly---LeuGlnGluAspGlnGluArgArgArgSerGlnGluGlnArgArgAspGln 647

Qy        958 TCTGCTGACATACAA-----------------------AATCATACATTAGAGACAGTA 993
Db        648 LysTrpArgTrpGlnLeuGluGluGluArgLysArgArgArgHisThrLeu--------- 664

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
Db        665 --------------------TyrAlaLysProAlaLeuGlnGluGlnLeuArgLysGlu 677

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
Db        678 ------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu--------- 692

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173
Db        693 ---------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArgGluGluGlu 709

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
Db        710 GlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGluLysArgArg 727

Qy       1234 GAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293

Db        728 ArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLysGluGluGln 747

Qy       1294 AATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT 1350
Db        748 LeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLysTyrArgGlu 767

Qy       1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db        768 GluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 780
```

RESULT 15
AAW03626
ID AAW03626 standard; peptide; 412 AA.
XX
AC AAW03626;
XX
DT 25-MAR-2003 (revised)
DT 17-DEC-1996 (first entry)
XX
DE Human thyrotropin GPR N-terminal sequence.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
XX
OS Homo sapiens.
XX
PN US5508384-A.
XX
PD 16-APR-1996.
XX
PF 09-SEP-1993; 93US-00118270.
XX
PR 10-SEP-1992; 92US-00943236.
XX
PA (UYNY) UNIV NEW YORK STATE.
XX
PI Schuster DI, Murphy RB;
XX
DR WPI; 1996-208785/21.
XX
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
PT treating schizophrenia.
XX
PS Disclosure; Fig 8B(2); 184pp; English.
XX
CC Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomegaloviral and other GPR proteins. The peptides AAW03578-
CC W03651 represent the N-terminal fragments of the above proteins. The
CC receptor proteins were used to design polypeptides, pref. based on the
CC transmembrane domains, for use in G-protein coupled receptor ligand
CC binding assays. The polypeptide fragments retain biological activity such
CC as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
CC AAW02747-W02999 for examples of polypeptide fragments). The polypeptide
CC fragments can be used in compositions for treating subjects suffering
CC from a pathology related to a GPR abnormality e.g. a psychotic disorder
CC such as schizophrenia. (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 412 AA;

Alignment Scores:

Pred. No.:	9.31e-19	Length:	412
Score:	362.50	Matches:	80
Percent Similarity:	51.73%	Conservative:	129
Best Local Similarity:	19.80%	Mismatches:	140
Query Match:	14.58%	Indels:	55
DB:	2	Gaps:	5

US-09-837-344-45 (1-1482) x AAW03626 (1-412)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||::::::   :::   |||:::|||        :::|||:::  :::|||:::
Db         33 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 52

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::   :::   |||::::::         :::|||:::  :::|||::::::   :::
Db         53 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 72

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                 |||:::|||         :::|||:::  :::|||::::::   :::   |||:::
Db         73 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 92

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||         :::|||:::  :::|||::::::   :::   |||:::|||
Db         93 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 112

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::|||:::   :::|||::::::   :::   |||::::::         :::|||
Db        113 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 132

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::   :::|||::::::   :::   |||:::|||         :::|||:::   :::
Db        133 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 152

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||::::::   :::   |||:::|||        :::|||:::  :::|||::::::
Db        153 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 172

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::   |||:::|||         :::|||:::  :::|||::::::   :::
Db        173 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 192

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::|||         :::|||:::  :::|||::::::   :::   |||:::|||
Db        193 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 212

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                  :::|||:::   :::|||::::::   :::   |||:::|||
Db        213 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 232

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::|||:::   :::|||::::::   :::   |||:::   :::   :::   ::::::
Db        233 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                    |||   ::::::|||
Db        253 GluGluGluGluGluGlu------------------------------------------ 259

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                          |||         |||   |||
Db        260 ---------------GluGluGluGluGluGluGlu------------------------ 266

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
                        ::::::                  :::   :::|||
Db        267 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 282

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::         |||                 |||         :::     :::
Db        283 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 302

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ||||||:::::::::         :::   |||::::::|||::::::
Db        303 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 322

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                 :::
Db        323 GluGlu------------------------------------------------------ 324

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                          |||:::|||         |||   ::::::         :::||||||
Db        325 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 340

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||::::::|||:::   :::|||   :::       |||   ::::::       |||:::
Db        341 GluGluGluGluGluGluGluGluGluGluTyrValPhePheGluGlu---GlnGlu 359

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::|||   ||||||      :::||||||::::::         :::   |||   |||
Db        360 AspGluIleIleGlyPheGlyGlnGluLeuLysAsn------ProGlnGluGluThrLeu 377

Qy       1201 GATGATTTAGAT 1212
                    |||
Db        378 GlnAlaPheAsp 381
```

```
RESULT 16
AAY96255
ID   AAY96255 standard; protein; 1162 AA.
XX
AC   AAY96255;
XX
DT   12-SEP-2003  (revised)
DT   11-SEP-2000  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;
KW   latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KW   Human herpes virus 8; HHV8; rhadino virus cis-acting element; RVCAE;
KW   Kaposi's sarcoma; primary effusion lymphoma; PEL;
KW   human immunodeficiency virus; HIV; multicentric Castleman's disease.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Domain          14. .17
FT                   /note= "nuclear localisation signal, NLS"
FT   Domain          64. .70
FT                   /note= "nuclear localisation signal, NLS"
FT   Region          320. .429
FT                   /note= "acidic repeat region"
FT   Region          430. .549
FT                   /note= "Gln, Glu, Pro-rich region"
FT   Region          550. .589
FT                   /note= "Gln, Glu, Pro, Arg-rich region"
FT   Region          590. .759
FT                   /note= "Gln, Glu, Asp-rich region"
FT   Region          760. .840
FT                   /note= "Gln, Glu-rich region"
XX
PN   WO200029626-A1.
XX
PD   25-MAY-2000.
XX
PF   19-NOV-1999;   99WO-US027508.
XX
PR   19-NOV-1998;   98US-00109422.
PR   21-APR-1999;   99US-00298568.
XX
PA   (KIEF/) KIEFF E D.
PA   (BALL/) BALLESTAS M E.
PA   (KAYE/) KAYE K M.
XX
PI   Kieff ED,  Ballestas ME,  Kaye KM;
XX
DR   WPI; 2000-387829/33.
DR   N-PSDB; AAA30290.
XX
PT   Treating or preventing a disease associated with rhodino virus infection
PT   in a mammal which includes Kaposi's Sarcoma and Primary Effusion
PT   Lymphoma.
XX
PS   Disclosure; Fig 7; 70pp; English.
XX
CC   The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC   (KSHV) latency-associated nuclear antigen (LANA). KSHV is also known as
```

```
CC   Human Herpes Virus 8 (HHV8) and belongs to the rhadino virus, or gamma-2
CC   herpes virus class. The LANA protein is necessary for the efficient
CC   persistence of rhadino virus DNA in mammalian cells. Persistent rhadino
CC   virus infection is implicated in a variety of diseases e.g. Kaposi's
CC   Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
CC   Castleman's disease. In addition, KS is a common malignancy in HIV
CC   patients. KSHV persists in host cells in a latent form. One of the few
CC   genes expressed from the latent viral DNA is LANA. LANA associates with
CC   both human chromosomes and with the rhadino virus cis-acting element
CC   (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC   "tied" to the host chromosomes. This allows the viral DNA to persist in
CC   the host cell. The present sequence may be used to screen and identify
CC   molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC   with the latency cycle of this virus. Potential antiviral treatments for
CC   the above mentioned diseases may therefore be based on LANA deregulation.
CC   (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:                1.92e-18        Length:          1162
Score:                    359.00          Matches:         94
Percent Similarity:       51.19%          Conservative:    121
Best Local Similarity:    22.38%          Mismatches:      117
Query Match:              14.44%          Indels:          88
DB:                       3               Gaps:            12

US-09-837-344-45 (1-1482) x AAY96255 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||::::::|||          ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                  :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      :::      :::|||              |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu-------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                 |||   |||
Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                       :::   :::|||      :::
Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                 |||:::         |||                  |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::      |||||||||   ::::::          :::   |||:::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGluGlnGluGlnGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::          |||   ::::::              |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                              :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              ::::::|||                  :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                              :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 17
AAY58500
ID   AAY58500 standard; protein; 1162 AA.
XX
AC   AAY58500;
XX
DT   06-AUG-2003  (revised)
DT   10-APR-2000  (first entry)
XX
DE   HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW   HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen; antigen.
XX
OS   Human herpesvirus 8.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 96
FT                   /label= unknown
XX
PN   WO9961909-A2.
XX
PD   02-DEC-1999.
XX
PF   26-MAY-1999;   99WO-US011407.
XX
PR   26-MAY-1998;   98US-0086695P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pau C;
XX
DR   WPI; 2000-097142/08.
XX
```

PT New methods and compositions for the detection of human herpesvirus.
XX
PS Claim 2; Page 59-62; 68pp; English.
XX
CC Sequences AAY58480-Y58532 represent immunogenic polypeptides derived from
CC human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC biological sample, and detecting the formation of a complex between the
CC peptide and the antibody. The presence of a peptide-antibody complex
CC indicates the presence of human herpesvirus 8. The detection of HHV8
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC HHV8-specific antibodies are useful therapeutically when for the passive
CC immunisation of a human against HHV8 infection, thereby reducing HHV8
CC related disease. The detection assays are highly specific, sensitive and
CC accurate. Early detection and treatment of Kaposi's sarcoma could
CC diminish the severity of symptoms related to AIDS and the sensitive
CC techniques could reduce erroneous characterisations of skin disorders.
CC Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC immunoblots and enzyme immunosassays lack the sensitivity and accuracy
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages of
CC the assays are that reproducible results are obtained and the method is
CC suitable for rapid throughput and screening of samples economically.
CC (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.92e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	3	Gaps:	12

US-09-837-344-45 (1-1482) x AAY58500 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||:::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||          :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::     |||:::|||                :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                 |||::::::      |||:::   |||::::::|||         ||||||:::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                   :::::::::   ::::::||||||   :::   :::|||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::             :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
```

```
                ||||||:::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                :::          :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
                :::      :::      :::|||             |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                    |||   |||
Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                    :::   :::|||      :::
Db        832 -----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                |||:::          |||                  |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                :::      |||||||||   ::::::          :::   |||:::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGlnGluGluGlnGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                :::::::   :::   |||::::::          |||   ::::::            |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
                |||:::::::::::::::|||:::|||                          :::
Db        897 GluGlnGluGluGluGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
                :::::::|||               :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln------------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                          :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937
```

RESULT 18
AAB62331
ID AAB62331 standard; protein; 1162 AA.
XX
AC AAB62331;
XX
DT 06-AUG-2003 (revised)
DT 29-JUN-2001 (first entry)
XX
DE Amino acid sequence of KSHV tethering protein LANA.
XX
KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW KSHV; latency-associated nuclear antigen; LANA.
XX
OS Human herpesvirus 8.
XX
PN WO200125484-A2.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US026908.

```
XX
PR   01-OCT-1999;   99US-00410399.
XX
PA   (UNMI ) UNIV MICHIGAN.
XX
PI   Robertson ES,  Cotter MA;
XX
DR   WPI; 2001-281736/29.
DR   N-PSDB; AAF82901.
XX
PT   A composition for use in gene therapy comprises an expression vector that
PT   includes a nucleic acid sequence encoding a nucleic acid binding protein.
XX
PS   Disclosure; Fig 9B; 60pp; English.
XX
CC   The invention provides a composition comprising nucleic acid, histone H1
CC   protein and expression vector operationally encoding a protein suitable
CC   for tethering the nucleic acid to the histone H1 protein, where the
CC   tethering protein is LANA. The composition is useful in aiding the
CC   retention of the viral DNA in the host cell. The viral vector encodes a
CC   protein suitable for tethering DNA to Histone H1. Methods for screening
CC   for compounds which are agonistic or antagonistic for the tethering of
CC   viral proteins to histone H1 and DNA binding sites are useful for
CC   developing the method of viral transfer. The composition has applications
CC   to gene therapy, including the treatment of multiple sclerosis,
CC   Parkinson's disease, Huntington disease and diabetes. The present
CC   sequence represents the amino acid sequence of the Kaposi's sarcoma
CC   associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC   antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-
CC   2003 to correct OS field.)
XX
SQ   Sequence 1162 AA;

Alignment Scores:
Pred. No.:                1.92e-18        Length:          1162
Score:                    359.00          Matches:         94
Percent Similarity:       51.19%          Conservative:    121
Best Local Similarity:    22.38%          Mismatches:      117
Query Match:              14.44%          Indels:          88
DB:                       4               Gaps:            12

US-09-837-344-45 (1-1482) x AAB62331 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||:::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::  :::||||||       :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||            :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                ||:::::::      |||:::   |||::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::

Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::       :::       :::|||            |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                  |||   |||
Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                        :::   :::|||      :::
Db        832 -----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                |||:::         |||                |||       ::::::
Db        840 GluGluGlnGluValGluGluGlnGlnGluValGluGluGlnGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::       |||||||||   ::::::         :::   |||::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGlnGluGluGlnGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::         |||   ::::::              |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                              :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              ::::::|||                  :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln------------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                           :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

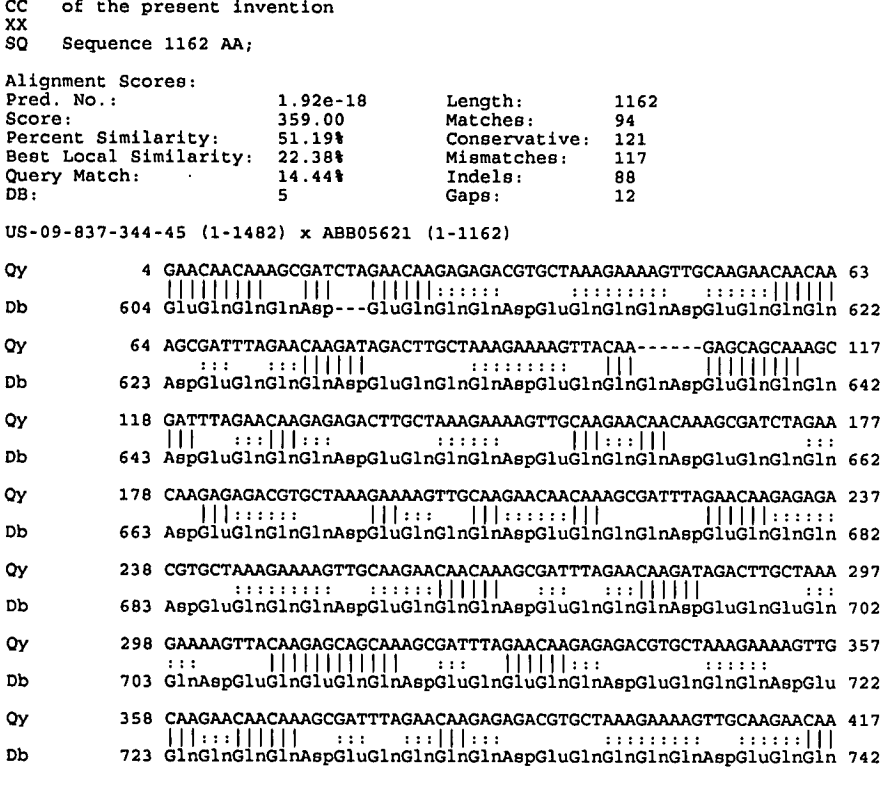
RESULT 19
ABB05621
ID   ABB05621 standard; protein; 1162 AA.
XX
AC   ABB05621;
XX
DT   25-APR-2002  (first entry)
XX
DE   Kaposi's sarcoma-associated herpesvirus LANA protein.
XX
KW   Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
KW   KSHV terminal repeat; rhadino virus cis acting element; episome;
KW   primary effusion lymphoma; latency-associated nuclear antigen;
KW   gene therapy; gene transfer.
XX
OS   Human herpesvirus 8.
XX
PN   US6322792-B1.
XX
```

PD 27-NOV-2001.
XX
PF 21-APR-1999; 99US-00298568.
XX
PR 19-NOV-1998; 98US-0109422P.
XX
PA (KIEF/) KIEFF E D.
XX
PI Kieff ED, Ballestas ME, Kaye KM;
XX
DR WPI; 2002-153769/20.
DR N-PSDB; ABA93487.
XX
PT System for episomal retention of plasmids in mammalian cells, useful in
PT gene therapy, comprises rhadinoviral LANA and RVCAE sequences.
XX
PS Disclosure; Fig 7; 27pp; English.
XX
CC The present invention describes a system (A) for maintaining a plasmid as
CC an episome in mammalian cells, comprising the rhadinoviral sequence LANA
CC (latency-associated nuclear antigen) of 3489 base pairs (see ABA93487,
CC S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC present in the plasmid. Also describes is a method for maintaining a
CC closed circular DNA in a cell by expressing (S1) in the cells and having
CC (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC particularly used in gene therapy (or other gene transfer applications)
CC that uses mammalian cells in which LANA is expressed. (A) improves
CC persistence of gene therapy vectors in cells. The present sequence
CC represents Kaposi's sarcoma-associated herpesvirus (KSHV, also called
CC human herpesvirus 8) LANA protein, which is used in the exemplification
CC of the present invention
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.92e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	5	Gaps:	12

US-09-837-344-45 (1-1482) x ABB05621 (1-1162)



RESULT 20
ADJ65096
ID ADJ65096 standard; protein; 1162 AA.
XX
AC ADJ65096;
XX
DT 20-MAY-2004 (first entry)
XX
DE HHV8 latency-associated nuclear antigen, LANA.
XX
KW HHV8; latency-associated nuclear antigen; LANA;
KW primary effusion lymphoma; PEL; virucide; gene therapy; KSHV;
KW Kaposi's sarcoma-associated herpesvirus; episome; RVCAE;
KW rhadinovirus cis-acting element; rhadinovirus infection.
XX
OS Human herpesvirus 8.
XX

PN US2004037847-A1.
XX
PD 26-FEB-2004.
XX
PF 28-JUN-2001; 2001US-00894273.
XX
PR 19-NOV-1998; 98US-0109422P.
PR 21-APR-1999; 99US-00298568.
XX
PA (KIEF/) KIEFF E D.
PA (BALL/) BALLESTAS M E.
PA (KAYE/) KAYE K M.
XX
PI Kieff ED, Ballestas ME, Kaye KM;
XX
DR WPI; 2004-191011/18.
DR N-PSDB; ADJ65095.
XX
PT Assays for compounds that modulate rhadino virus LANA action in trans on
PT a unit of rhadino virus DNA to mediate efficient episome persistence,
PT comprises detecting whether the compound inhibits expression of a gene
PT resident on that episome.
XX
PS Disclosure; SEQ ID NO 2; 28pp; English.
XX
CC The invention relates to an assay for determining whether a compound
CC modulates LANA (latency-associated nuclear antigen) from KSHV (Kaposi's
CC sarcoma-associated herpesvirus, also known as HHV8, human herpesvirus 8)
CC mediated persistence in a mammalian cell of a DNA episome having RVCAE
CC (rhadinovirus cis-acting element) by detecting whether the compound
CC inhibits expression of a gene resident on that episome. Also included are
CC maintaining a closed circular DNA in a mammalian cell in which LANA is
CC expressed (by using RVCAE as the cis-acting origin and maintenance
CC sequence) and maintaining a plasmid as an episome in a mammalian cell
CC (comprising: (a) Expressing LANA in the cell; and (b) having RVCAE
CC resident on the plasmid). Compounds identified as interfering with the
CC interaction between LANA and RVCAE are useful in treating or preventing
CC rhadino virus infection (e.g. primary effusion lymphoma (PEL)). LANA acts
CC in trans on a 0.8 Kb KSHV TR (terminal repeat) unit to mediate efficient
CC episome persistence. The present sequence represents KSHV LANA.
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.92e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	8	Gaps:	12

US-09-837-344-45 (1-1482) x ADJ65096 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                ||:::::::      |||:::   |||::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::           :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      :::      :::|||                |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                  |||   |||
Db        828 ----------------------------GlnGluGlnGlu---------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                        :::   :::|||      :::
Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                 |||:::         |||                |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::      |||||||||   ::::::         :::   |||::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::         |||   ::::::                |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                              :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu---------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              ::::::|||               :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                              :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937
```

RESULT 21
ADV68154
ID ADV68154 standard; protein; 1162 AA.
XX
AC ADV68154;
XX
DT 24-FEB-2005 (first entry)
XX
DE Kaposi's sarcoma-associated herpesvirus LANA protein.

XX
KW screening; cytostatic; virucide; gene therapy;
KW latency-associated nuclear antigen; tumor; viral infection;
KW herpesvirus type 8 infection.
XX
OS Human herpesvirus 8.
XX
PN US2004248081-A1.
XX
PD 09-DEC-2004.
XX
PF 11-JUL-2002; 2002US-00194046.
XX
PR 01-OCT-1999; 99US-00410399.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Robertson ES;
XX
DR WPI; 2005-038492/04.
DR N-PSDB; ADV68153.
XX
PT New composition having a latency-associated nuclear antigen (LANA)
PT nucleic acid, useful for detecting compounds for treating tumors and
PT viral infections.
XX
PS Disclosure; SEQ ID NO 2; 51pp; English.
XX
CC The invention relates to a composition comprising a latency-associated
CC nuclear antigen (LANA) nucleic acid having any of 14 fully defined
CC sequences of 10-61 bp given in the specification (ADV68159-ADV68172). The
CC methods and compositions of the present invention are useful for
CC detecting compounds that are agonistic or antagonistic for the binding of
CC viral genetic material to genomic host DNA, in particular for detecting
CC compounds for treating tumor and viral infections. This sequence
CC corresponds to the LANA protein from the Kaposi's sarcoma-associated
CC herpesvirus (also called human herpesvirus 8). The LANA protein tethers
CC the viral DNA to the human chromosomal structural protein histone H1.
XX
SQ Sequence 1162 AA;

Alignment Scores:

Pred. No.:	1.92e-18	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	9	Gaps:	12

US-09-837-344-45 (1-1482) x ADV68154 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                :::   :::||||||          :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||::::::|||          ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
```

```
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::          :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                  ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      :::      :::|||                |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                 |||   |||
Db        828 ----------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                       :::   :::|||      :::
Db        832 ------------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                 |||:::         |||                |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::      |||||||||   ::::::          :::   |||:::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::          |||   ::::::              |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                                :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              :::::::|||                  :::|||||||||:::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                                  :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937
```

RESULT 22
AAB44672
ID AAB44672 standard; protein; 207 AA.
XX
AC AAB44672;
XX
DT 12-FEB-2001 (first entry)
XX
DE Gene 33 human secreted protein homologous amino acid sequence #137.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

```
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW   food additive; preservative; chromosome identification; cancer;
KW   immune disorder; cardiovascular disorder; neurological disease;
KW   wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200058339-A2.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US007440.
XX
PR   26-MAR-1999;   99US-0126503P.
PR   17-DEC-1999;   99US-0172409P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594637/56.
XX
PT   Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT   the prevention, treatment and diagnosis of cancer, immune disorders,
PT   cardiovascular disorders and neurological diseases.
XX
PS   Disclosure; Page 399-400; 410pp; English.
XX
CC   The polynucleotide sequences given in AAC79681 to AAC79730 encode the
CC   human secreted proteins given in AAB44596 to AAB44645. AAB44646 to
CC   AAB44693 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
CC   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC   cardiant. The polynucleotides and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC   can also be used as a food additive or preservative to increase or
CC   decrease storage capabilities. The polynucleotides are useful for
CC   chromosome identification. They are also useful as probes for diagnosing
CC   a disorder related to the female reproductive system, particularly breast
CC   and/or ovary cancer. They are also useful in the gene therapy of breast
CC   and ovarian cancer. The nucleic acids, protein, antibodies, agonists and
CC   antagonists from the present invention are useful in the diagnosis,
CC   treatment and prevention of: cancer; immune disorders; cardiovascular
CC   disorders; wound healing; neurological diseases; and infectious diseases.
CC   AAC79672 to AAC79680 and AAB44595 represent sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 207 AA;

Alignment Scores:
Pred. No.:              4.51e-17        Length:         207
Score:                  339.50          Matches:        75
Percent Similarity:     72.25%          Conservative:   89
Best Local Similarity:  33.04%          Mismatches:     34
Query Match:            13.65%          Indels:         29
DB:                     3               Gaps:           12

US-09-837-344-45 (1-1482) x AAB44672 (1-207)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::|||:::      |||||||||
Db          1 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGluGlnGlnAspGluGlnGln 19

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                  |||   :::|||   :::      :::|||:::   :::|||||||:::
Db         20 GlnAspGluGlnGlnGlnGlnAspGluGlnGluGlnGlnGluGluGlnGluGlnGlnGlu 39

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||||||||:::       :::|||:::   ||||||||||   :::   |||||||||
Db         40 GluGlnGluGln---GlnGluGluGlnGluGlnGluGlnGlnGluGluGlnGluGlnGlu 58

Qy        184 AGACGTGCTAAAGAA---AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
                    :::|||   :::|||:::||||||:::   :::||||||::::::
Db         59 LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 76

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::::::::|||:::||||||:::   :::|||||||:::            ::::::
Db         77 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln---------GluGln 92

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::|||:::||||||:::   :::||||||::::::         :::::::::|||:::
Db         93 GluLeuGluGluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGlu 109

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||:::   :::|||||||::::::         :::::::::|||:::||||||:::
Db        110 GluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGlu 126

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::||||||::::::         :::::::::|||:::||||||:::   :::|||
Db        127 GlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeu 143

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||::::::         :::::::::|||:::||||||:::   :::||||||::::::
Db        144 GluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln 160

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                    :::::::::|||:::||||||:::   :::||||||::::::
Db        161 ---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln--------- 174

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::::::::|||:::||||||:::::::::||||||::::::      :::   ::::::
Db        175 GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------GluGlnGluGln 192

Qy        661 AATTTAGAAAGAAAAAAGGAA 681
                  ||||||      :::|||
Db        193 GluLeuGluGluValGluGlu 199

RESULT 23
AAR70491
ID   AAR70491 standard; protein; 562 AA.
XX
AC   AAR70491;
XX
DT   27-AUG-2003  (revised)
DT   19-DEC-1995  (first entry)
XX
DE   Leucocytozoan protozoa structural protein epitope.
XX
KW   leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
KW   leucocytozoanosis; treatment.
XX
OS   Leucocytozoon.
XX
PN   JP07089995-A.
XX
PD   04-APR-1995.
XX
PF   10-SEP-1993;   93JP-00226078.
XX
PR   10-SEP-1993;   93JP-00226078.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (NISS-) NISSEIKEN KK.
XX
DR   WPI; 1995-167252/22.
DR   N-PSDB; AAQ87587.
XX
PT   Immune inducing polypeptide against Leucocytozoan protozoa - useful in
PT   production of vaccines for treatment of leucocytozoanosis in fowl.
```

XX
PS Claim 1; Page 12-14; 20pp; Japanese.
XX
CC AAR70491-93 are polypeptides having a whole or partial epitope of a
CC structural protein of Leucocytozoan protozoa encoded by AAQ87587-89. The
CC polypeptides and DNA encoding them are useful in the production of
CC vaccines for the treatment of leucocytozoanosis of fowl. (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 562 AA;

Alignment Scores:

Pred. No.:	1.1e-16	Length:	562
Score:	335.00	Matches:	118
Percent Similarity:	51.07%	Conservative:	120
Best Local Similarity:	25.32%	Mismatches:	146
Query Match:	13.47%	Indels:	82
DB:	2	Gaps:	22

US-09-837-344-45 (1-1482) x AAR70491 (1-562)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::::::::   :::   |||||||||   :::   :::|||   ::::::||||||
Db         44 GluGlnGluGluGluGluGlnGluGlnGluGluGlnGluGlnGluIleValGluGluGln 63

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::   |||   |||:::           :::|||:::   :::|||::::::   :::
Db         64 GluGlnAspGluGluGluGln---------GluGluGluAspGluGluGluGluGluGlu 80

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                 |||:::|||         :::|||:::   :::|||::::::   :::   |||:::
Db         81 LysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              :::         :::|||:::   :::||||||      |||   |||:::|||
Db        101 GlnAspGluGluGluGluGluGluGluGluGlnAspGluAspGluGluGluGluGluAsp 120

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---CTTGCTAAA 297
                 :::|||:::   :::|||:::|||::::::   ||||||:::
Db        121 GluGluGluGluGluGluGluGluGlnAsnGluGluGluGlnAsnGluAspGluGlnAsn 140

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              |||       |||   ::::::   :::   |||:::|||         :::|||:::
Db        141 GluAspGluGlnAsnGluGluGluGluGluGluGluGluGluGlnGlnGluGlnAsp 160

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              :::|||:::|||   :::   ||||||:::         :::|||:::   ||||||:::
Db        161 GluGluGluGlnAspGluGluGluGlnAspGluGluGluGluGluGluGlnGluGlu 180

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCT--------------------------AAA 450
              |||   :::   ::::::|||      |||                          |||
Db        181 GlnAspGluGluGlnGluGluValTyrAlaGluLysGluAsnGluAspGluGluLysLys 200

Qy        451 GAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGTGCTAAA 501
              ||||||   :::||||||:::   :::         :::|||:::|||:::      :::
Db        201 GluLysGluGluGluGlnGluAspGluLysIleTyrValGluLysGluLysAspGluGlu 220

Qy        502 ---------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                      ||||||   ||||||         :::||||||:::|||      :::
Db        221 ValTyrValGluLysGluGlnGluHisPheLysGluLeuGluLysGlu---ValLeuGlu 239

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||::::::   :::       |||   ::::::|||   :::   |||:::
Db        240 GluLysValGluSerGluHisLysAspAspLysLysGluThrLysIleLysLysHisGlu 259

Qy        613 CAAGAGCAGCAAAGAGATTTA------------GAACAAAGGAAGGCTGATACGAAAAAA 660
                 |||::::::|||::::::            |||       :::   |||   ||||||
Db        260 ValGluGluGluArgGluIleGluIleLysGlyGluSerGluGluIleAspGluLysLys 279

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              :::   |||   |||::::::::::|||   :::       :::
Db        280 Ser---GluValLysGluGlnAsnGlyGlnValGlnGlyLys----------------- 292
```

```
Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA------CCA 774
                                                 ||||||:::      ||||||      |||
Db        293 --------------------------ThrAsnGluGlnLysGluTyrIleGlnAspPro 303

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCT 834
                    ||||||:::   :::                         ::::::
Db        304 IleTrpSerSerIleLeuArgTyrMetIleAsnLysArgThrGluThr------------ 319

Qy        835 ATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT------GAAGGACGAAGG 888
                  ||||||||||||:::::::::             |||       |||   :::
Db        320 ------GluLysThrAsnGlnGlnAsnTyrAsnIleLysValThrAsnGluSerLysSer 337

Qy        889 GATATACATAAAGGA---------CATCTTGAAGAAAAGAAAGATGGTTCAATA---AAA 936
              :::|||:::                     :::|||   |||:::   |||:::   |||
Db        338 GluIleAsnSerThrThrAsnIlePheGluAspGluAsnLysAsnSerSerLeuLysLys 357

Qy        937 CCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
              ||||||      |||                                 :::|||   |||
Db        358 ProGluLeuAsnGluAlaThrArgGlyArgAsnSerGlyProValValGluGluVal--- 376

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA--- 1053
                  |||   :::       :::         |||:::|||:::       |||
Db        377 ---------ValGluGluThrProVal------ValGluGluGluMetValGluGluThr 391

Qy       1054 ------TATGACGATTCATTAATAGATGAA------------GAAGAAGAT--------- 1086
                        ::::::   :::|||:::|||           ||||||:::
Db        392 ProValValGluGluGluMetIleGluGluThrProValValGluGluGluMetIleGlu 411

Qy       1087 --------------GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
                           :::||||||:::   |||       |||||||||:::   |||
Db        412 GluThrProValValGluGluAspValValGluGluThrProIleValGluGluAsp--- 430

Qy       1132 TTCCAAGATGAAGAAAACATA------GGAATTTATAAAGAACTA------------GAA 1173
                    :::|||   |||:::         :::   :::|||                |||
Db        431 ValValGluGluThrAsnValValGluGluValValGluGluThrProValValGluGlu 450

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
              ::::::||||||:::          ::::::::::::   |||
Db        451 GluMetIleGluGluThrProValValGluGluValValGluGluThrProValValGlu 470

Qy       1234 GAAGAATTATCTGAAGAA 1251
              ||||||:::   ||||||
Db        471 GluGluValValGluGlu 476
```

RESULT 24
AAB18165
ID AAB18165 standard; protein; 1192 AA.
XX
AC AAB18165;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:22.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US026796.
XX
PR 05-NOV-1998; 98US-0107131P.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;

```
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection.
XX
PS   Disclosure; Page 60-63; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC   (II) are useful for the development of vaccines against P. falciparum
CC   infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC   immunogens comprising the sequences of (I), are useful in the detection
CC   of infection with P. falciparum. Furthermore, (I) (especially when they
CC   are rifins or secreted or membrane proteins) can aid the identification
CC   of drugs to treat or prevent P. falciparum infection, or they can be used
CC   to identify drug resistance in P. falciparum. Sequencing of the
CC   Plasmodium chromosome 2 and the subsequent identification of proteins
CC   encoded by it will help to expand our understanding of parasite biology,
CC   a process hampered by the complexity of the parasitic lifecycle, and
CC   provide new targets for vaccine and drug development. Parasite resistance
CC   to drugs and mosquito resistance to insecticides have led to a resurgence
CC   of malaria in many parts of the world, and there is a pressing need for
CC   vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC   represent nucleotide and protein sequences given in the present
CC   invention, but which are not specifically mentioned within the
CC   specification
XX
SQ   Sequence 1192 AA;

Alignment Scores:
Pred. No.:                1.31e-16      Length:          1192
Score:                    334.50        Matches:         112
Percent Similarity:       46.85%        Conservative:    126
Best Local Similarity:    22.05%        Mismatches:      193
Query Match:              13.45%        Indels:          77
DB:                       3             Gaps:            16

US-09-837-344-45 (1-1482) x AAB18165 (1-1192)

Qy         10 CAAAGCGATCTAGAACAAGAGAGA---------------------------------CGT 36
              ||||||      :::|||||||||                                   :::
Db        296 GlnSerLysTyrLysGlnGluArgIleGluIleLeuAspAspAsnGlyLysGluLeuLys 315

Qy         37 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA---------GATAGACTT 87
              :::   :::   ::::::|||:::::      :::|||:::                :::
Db        316 SerHisLysAsnIleLysGluGluLysGlyGlyIleGluLysThrAspThrThrAsnIle 335

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 147
              |||      |||:::::::::::::::   :::   :::   |||:::      ::::::
Db        336 AlaAspIleLysIleLysLysGluGluArgGluThrLysAspGluLys------GluLys 353

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207
                 :::|||:::         |||::::::      :::      :::|||
Db        354 AsnIleGlnGlnLeuValLysAspValGlnLeuIleLysValGlyGluGluThrLysAsp 373

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
              :::::::::   |||   |||              ::::::      |||:::
Db        374 AspGluLysGluAspLysGluGlyThrAspAspGluGluAspThrAspAspGluGluAsp 393

Qy        268 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 327
              :::|||   |||:::|||          :::|||          ::::::::::::
Db        394 ThrAspAspGluGluAspThrAspAspGluGluAspThrSerAspGluGluThrThrGly 413

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::||||||   :::   :::   ::::::   |||:::::::::::
Db        414 AspGlnGluAsnLysGluGluThrGluValAspGluLysLysThrGlu------------ 429

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
                 :::|||:::|||:::|||:::|||   :::   :::   |||:::::::::

Db        430 ---LysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGluGlu 448

Qy        448 AAAGAAAAG---TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 504
                 ||||||      :::|||   :::   |||   |||:::      :::   ||||||
Db        449 SerGluLysAspLysGluGluSerGluLysAspLysGluGluSerGluLysAspLysGlu 468

Qy        505 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 564
              |||   :::|||   :::          |||   |||:::      :::         :::
Db        469 LysThrGluGluAspGluGluLysThrGluAspGluLysGlyThrGluValTyrLysLys 488

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAA 624
              |||         |||   ::::::|||:::            :::      :::   :::
Db        489 GluThrAspValAspGluLysLysGluLysGlyGluTyrGlyGluGlyThrAspAspGlu 508

Qy        625 AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACAT 684
                 |||   |||:::   :::   |||   :::      :::|||   ||||||
Db        509 GluAspLysGluLysGluGluAspAspGluGluThrLysValGluGluLysLysThrGlu 528

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCA 744
              |||
Db        529 LysAsp---------------------------------------------------- 530

Qy        745 TCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA 804
                    |||   |||                  :::         :::|||
Db        531 ---------GluGluGly-----------------ThrAspTyrGluGluAspThrAsp 541

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA---GAATCT 861
                 :::   :::|||   :::                  :::||||||   |||   |||
Db        542 AspSerAspLysAspGluGluThrLysValGluGluLysLysThrGluArgAspGluGlu 561

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA 921
                 |||         |||         ::::::   |||         |||:::   :::
Db        562 GluThrGluGluAspGluLysGluThrGluValGluLysLysLysThrGluLysAspGlu 581

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                 :::|||:::         |||      :::|||         |||::::::
Db        582 GluGlyThrAspTyrGluGluAspThrAspAspSerAspLysAspValGluThrGluVal 601

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAG 1035
                 ||||||            |||      |||   :::   ::::::         :::
Db        602 GluGluThr------------AspAlaGluAspLysGluGluAsnGluGluGlyThrAsp 617

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT---GATGAA 1092
                 |||||               :::      |||   |||:::|||||||||   :::|||
Db        618 AspGluGluAspLysValGluGluThrAspLeuAspAspGlnGluGluAspGlyGluGlu 637

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
                 |||   ::::::   |||         :::   |||:::   :::|||::::::
Db        638 AspLysGluAspAspLysGluLysAspLysGluAspAspLysGluAspAspLysGlu--- 656

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA---------------------AAT 1191
                    |||:::   ||||||      ||||||                        :::
Db        657 ---------LysAspLysGluAspAspLysGluLysTyrLysGluAspAspLysGluAsp 673

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
                 :::      ||||||   ::::::      |||   :::   |||:::      |||:::
Db        674 AspLysGluAspAspLysGluLysAspLysGluAspAsnLysGluLysAspLysGluAsp 693

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
                 :::|||   |||:::   :::|||   ||||||          |||   :::|||
Db        694 AsnLysGluLysAspLysGluAspAspLysGluLysAsp---------LysGluAspAsp 710

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
                 |||         |||      |||      :::   ||||||:::   |||||||||
Db        711 LysGluLysAspLysGluAspAsnLys------GluLysAspLysGluAspAsnLysGlu 728

Qy       1372 AAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                 |||:::|||                  |||      |||      :::   :::
Db        729 LysAspLysGluAspAspLysGluLysHis-----------AspLysHisValArgArg 744

Qy       1432 ATCGTGGATGAGTTATCTGAAGAT 1455
                 |||      ::::::   :::|||
Db        745 IleLysLysLysMetLysAspAsp 752
```

```
RESULT 25
ADP29884
ID   ADP29884 standard; protein; 2990 AA.
XX
AC   ADP29884;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human secreted protein SEQ ID #651.
XX
KW   Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
KW   cancer; inflammatory; immune; human secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO2004035732-A2.
XX
PD   29-APR-2004.
XX
PF   28-AUG-2003; 2003WO-US026780.
XX
PR   29-AUG-2002; 2002US-0406576P.
PR   29-AUG-2002; 2002US-0406579P.
PR   29-AUG-2002; 2002US-0406585P.
PR   29-AUG-2002; 2002US-0406588P.
PR   29-AUG-2002; 2002US-0406608P.
PR   29-AUG-2002; 2002US-0406611P.
PR   29-AUG-2002; 2002US-0406612P.
PR   29-AUG-2002; 2002US-0406616P.
PR   29-AUG-2002; 2002US-0406640P.
PR   29-AUG-2002; 2002US-0406642P.
PR   29-AUG-2002; 2002US-0406646P.
PR   29-AUG-2002; 2002US-0406653P.
PR   29-AUG-2002; 2002US-0406655P.
PR   29-AUG-2002; 2002US-0406666P.
PR   17-SEP-2002; 2002US-0410946P.
PR   17-SEP-2002; 2002US-0410947P.
PR   17-SEP-2002; 2002US-0410948P.
PR   17-SEP-2002; 2002US-0410949P.
PR   17-SEP-2002; 2002US-0410953P.
PR   17-SEP-2002; 2002US-0410957P.
PR   17-SEP-2002; 2002US-0410958P.
PR   17-SEP-2002; 2002US-0410959P.
PR   17-SEP-2002; 2002US-0410960P.
PR   17-SEP-2002; 2002US-0410961P.
PR   17-SEP-2002; 2002US-0410962P.
PR   17-SEP-2002; 2002US-0411019P.
PR   17-SEP-2002; 2002US-0411022P.
PR   17-SEP-2002; 2002US-0411023P.
PR   17-SEP-2002; 2002US-0411024P.
PR   17-SEP-2002; 2002US-0411032P.
PR   17-SEP-2002; 2002US-0411035P.
PR   17-SEP-2002; 2002US-0411037P.
PR   17-SEP-2002; 2002US-0411041P.
PR   17-SEP-2002; 2002US-0411045P.
PR   17-SEP-2002; 2002US-0411046P.
PR   17-SEP-2002; 2002US-0411048P.
PR   17-SEP-2002; 2002US-0411052P.
PR   17-SEP-2002; 2002US-0411055P.
PR   17-SEP-2002; 2002US-0411073P.
PR   17-SEP-2002; 2002US-0411082P.
PR   17-SEP-2002; 2002US-0411101P.
PR   17-SEP-2002; 2002US-0411111P.
PR   18-APR-2003; 2003US-0463700P.
PR   18-APR-2003; 2003US-0463708P.
PR   18-APR-2003; 2003US-0463716P.
PR   18-APR-2003; 2003US-0463732P.
PR   02-MAY-2003; 2003US-0467199P.
PR   02-MAY-2003; 2003US-0467201P.
PR   02-MAY-2003; 2003US-0467203P.
PR   02-MAY-2003; 2003US-0467230P.
PR   19-MAY-2003; 2003US-0471306P.
PR   19-MAY-2003; 2003US-0471336P.
PR   22-MAY-2003; 2003US-0472420P.
PR   22-MAY-2003; 2003US-0472430P.
PR   09-JUN-2003; 2003US-0476609P.
PR   09-JUN-2003; 2003US-0476641P.
PR   08-JUL-2003; 2003US-0485218P.
PR   08-JUL-2003; 2003US-0485223P.
PR   08-JUL-2003; 2003US-0485224P.
PR   08-JUL-2003; 2003US-0485325P.
PR   14-JUL-2003; 2003US-0486446P.
PR   14-JUL-2003; 2003US-0486480P.
PR   15-JUL-2003; 2003US-0486891P.
PR   15-JUL-2003; 2003US-0486960P.
PR   08-AUG-2003; 2003US-0493341P.
PR   08-AUG-2003; 2003US-0493370P.
PR   08-AUG-2003; 2003US-0493573P.
PR   08-AUG-2003; 2003US-0493577P.
XX
PA   (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI   Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI   Halenbeck RF, Huang MM, Kothakota S, Haishan L, Linnemann T;
PI   Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
XX
DR   WPI; 2004-348438/32.
XX
PT   New nucleic acid molecule for diagnosing, preventing or treating diseases
PT   such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT   genetic, bacterial and viral diseases.
XX
PS   Claim 1; SEQ ID NO 1882; 428pp; English.
XX
CC   The present invention relates to an isolated nucleic acid molecule
CC   encoding a polypeptide which is believed to be cytostatic,
CC   antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC   composition and methods are useful for diagnosing, preventing and
CC   treating diseases such as proliferative (e.g. cancer), inflammatory,
CC   immune, metabolic, genetic, bacterial and viral diseases. The present
CC   sequence represents a human secreted protein. The present sequence is
CC   available on WIPOWEB and is not in the specification.
XX
SQ   Sequence 2990 AA;

Alignment Scores:
Pred. No.:              2.92e-16      Length:          2990
Score:                  330.50        Matches:         156
Percent Similarity:     46.43%        Conservative:    91
Best Local Similarity:  29.32%        Mismatches:      128
Query Match:            13.29%        Indels:          157
DB:                     8             Gaps:            34

US-09-837-344-45 (1-1482) x ADP29884 (1-2990)

Qy          1 CAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAA---------AAG 48
              :::||||||:::       |||   |||||||||   |||   :::|||         :::
Db       1490 ArgGluGlnGluGluArgLeuArgGluGlnGluGluArgGlnArgGluGlnGluAspArg 1509

Qy         49 TTGCAAGAACAACAAAGCGATTTA------GAACAAGATAGACTTGCT------AAAGAA 96
              |||   ||||||:::     |||      :::::::::||||||         :::|||
Db       1510 LeuHisGluGlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluGlu 1529

Qy         97 AAGTTACAAGAGCAGCAAAGCGATTTA------GAACAAGAGAGACTTGCT------AAA 144
              :::|||   ||||||:::       |||         :::|||||||||         :::
Db       1530 ArgLeuCysGluGlnGluGluArgLeuArgGluHisGluGluArgLeuCysGluGlnGlu 1549

Qy        145 GAAAAGTTGCAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAA------ 195
              |||:::|||:::||||||:::       |||   |||||||||   |||   :::
Db       1550 GluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluGluArgLeuArgGluHis 1569

Qy        196 ---GAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGT--------- 240
                 |||:::|||   ||||||:::     |||   |||||||||   |||
Db       1570 GluGluArgLeuCysGluGlnGluGluArgLeuCysGluGlnGluGluArgLeuHisGlu 1589
```

```
Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGATAGACTTGCTAAA 297
Db      1590 GlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGlu----------- 1605

Qy       298 GAAAAGTTACAAGAGCAGCAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA------ 348
Db      1606 GluArgLeuArgGluHisGluGluArgLeuCysGluGlnGluGluArgLeuArgGluHis 1625

Qy       349 ---GAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGT--------- 393
Db      1626 GluGluArgLeuCysGluGlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGlu 1645

Qy       394 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACTTGCTAAA 450
Db      1646 GlnGluGluArgLeuArgGluGlnGluGluArgLeuCysGluGlnGluLysLeuProGly 1665

Qy       451 GAAAAG------TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 504
Db      1666 GlnGluArgLeuLeuGluGluValGluLysLeuLeuGluGlnGluArgArgGlnGluGlu 1685

Qy       505 AAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAACAA 537
Db      1686 GlnGluArgLeuLeuGluArgGluArgLeuLeuGluGluValGluLysLeuLeuGluGln 1705

Qy       538 GAGAGACGTGCTAAAGAAAAG--------------------------TTGCAAGAACAA 570
Db      1706 GluArgGlnGlnGluGluGlnGluArgLeuLeuGluArgGluArgLeuLeuGluGluVal 1725

Qy       571 CAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAGCAGCAA 624
Db      1726 GluLysLeuLeuGluGlnGluArgArgGlnGluGluGlnGluArgLeuLeuGluArgGlu 1745

Qy       625 AGA---------GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db      1746 ArgLeuLeuAspGluValGluGluLeuLeuAspGluThrLeuGlnGluLeuGluArgLeu 1765

Qy       676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
Db      1766 ArgGlu----------------------------------------------------Leu 1768

Qy       736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
Db      1769 GluArgLeuArgGluLeuGluArgMetLeuGluLeuGlyTrpGluAlaLeuTyrGluGln 1788

Qy       796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA 855
Db      1789 ------ArgAlaGluProArg--------------SerGlyPheGluGluLeuAsnAsn 1801

Qy       856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
Db      1802 GluAsnLysSerThr--------------LeuGlnLeuGluGlnGlnValLysGluLeu 1816

Qy       916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
Db      1817 LysLysSerGlyGluLeuLysGluThrValThrSerAspProSerLysLysMet------ 1834

Qy       976 CATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAG 1035
Db      1835 ------------------------------------TrpGluProIleValPheLys 1841

Qy      1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT--------- 1086
Db      1842 GluLysLeuThrMetLys---ThrAspSerLeuMetGluGluLysLeuGluCysSerLeu 1860

Qy      1086 ------------------------------------------------------------ 1086

Db      1861 TrpCysCysLeuSerAsnProSerIleProGlyArgCysCysValLeuGluArgArgIle 1880

Qy      1087 --------------GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
Db      1881 ValProTrpMetGlnGlnGluSerTyrSerSerSerSerProIleTrpSerValAspSer 1900

Qy      1132 TTCCAAGATGAAGAAAACATA---GGAATTTATAAAGAACTAGAAGATTTGATAGAG--- 1185

                                                                          
Db      1901 ------AspGluProAsnLeuThrSerValLeuGluArgLeuGluAspThrLysGluAsn 1918

Qy      1186 ------------------------------------AAAAATGAAAATTTAGATGAT 1206
Db      1919 SerSerProLeuPheAsnCysProProProThrThrSerLysThrGluIleLeuAspSer 1938

Qy      1207 TTAGAT--------------------GAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
Db      1939 CysAsnCysSerProSerGluSerValArgGlyAlaGluLysGlyAspGlnGluLeuLeu 1958

Qy      1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
Db      1959 ---AlaAsnIleLysGlnAlaGluLysHisGluLys 1969
```

RESULT 26
ABB58657
ID ABB58657 standard; protein; 2274 AA.
XX
AC ABB58657;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2763.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02760.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 2763; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2274 AA;

Alignment Scores:

Pred. No.:	3.08e-16	Length:	2274
Score:	330.00	Matches:	129
Percent Similarity:	43.83%	Conservative:	134
Best Local Similarity:	21.50%	Mismatches:	177
Query Match:	13.27%	Indels:	160
DB:	4	Gaps:	24

US-09-837-344-45 (1-1482) x ABB58657 (1-2274)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTG 51
              :::|||::::::      |||      ||||||     |||         ||||||:::
Db       1385 LysGluArgGluGluSerLeuArgMetGluArgGluGluLysLeuLysGluGluLysIle 1404

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
              :::|||::::::           :::::::::|||   :::||||||:::::::|||:::
Db       1405 LysGluLysGlu---------ArgGluGluLysLeuLysGluGluLysIleLysGluLys 1421

Qy        112 CAAAGCGAT------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
              |||   :::      |||:::::|||:::|||   :::::::::   :::|||:::
Db       1422 GlnArgGluGluLysLeuLysGluGluLysLeuLysGluLysGluArgGluGluArgMet 1441

Qy        166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA--------- 216
                 :::   |||:::|||   :::|||||||||||||   :::|||::::::
Db       1442 LysGluLysGluArgGluGluLysAlaLysGluLysGlnArgGluGluLysLeuArgGlu 1461

Qy        217 ---------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                      :::   |||:::|||   :::   ||||||||||||:::|||::::::
Db       1462 GluLysIleLysGluLysGluArgGluGluLysLeuLysGluLysLeuArgGluGluLys 1481

Qy        268 ---AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
                 :::   |||::::::      :::::::::   :::|||:::     :::
Db       1482 IleLysGluLysGluLysGluGluLysLeuArgLysGluArgGluGluLysMetArgGlu 1501

Qy        325 TTAGAACAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAACAA 369
                 |||:::|||   :::   |||                ||||||::::::|||::::::
Db       1502 LysGluArgGluGluLysIleLysGluLysGluArgValGluLysIleLysGluLysGlu 1521

Qy        370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                 :::   :::   ::::::      :::||||||||||:::|||::::::      |||
Db       1522 ArgGluGluLysLeuLysLysGluLysGluGluLysLeuLysGluLysGluGluLeuLeu 1541

Qy        430 GAA-------------------------CAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::                         :::|||:::|||   :::::::::   |||
Db       1542 LysLysLysGluLysGluGluLysGluArgGluGluLysLeuLysGluLysGluArgGln 1561

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||:::      :::   |||:::|||   :::   |||   :::   :::|||:::|||
Db       1562 GluLysLeuLysGluLysGluArgGluGluLysLeuLysArgGluThrGluGluArgGln 1581

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGAT 579
                 :::   |||:::|||   :::   |||||||||   :::   |||:::      |||
Db       1582 ArgGluLysGluArgGluGluLysLeuLysGluLysGluArgAlaGluLysLeuLysAsp 1601

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAGCAGCAAAGAGAT 630
              ||||||:::|||   :::   |||||||||         |||:::|||::::::::::::
Db       1602 LeuGluLysGluValLysLeuLysGluLysGluGluGlnLeuLysGluLysGluLysGlu 1621

Qy        631 TTA---------GAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              |||         |||:::::::::   |||      |||      |||:::      |||
Db       1622 LeuLysLeuLysGluLysLysGluLysAspLysValLysGluLysGluLysSerLeuGlu 1641

Qy        682 CATGGAGATATATTA-----------------------------------GCAGAGGAT 705
                    :::|||                                     ||||||
Db       1642 SerGluLysLeuLeuIleSerAlaThrValSerAsnProTrpArgArgValValGluAsp 1661

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATAGAA-------------------------- 738
                    :::|||      ||||||::::::
Db       1662 ThrProProLysLeu------ProAlaValGlnAspTyrProSerLeuGlyLysLysPro 1679

Qy        739 ---------CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                      |||   :::                   :::|||      ::::::
Db       1680 ThrLysAlaSerProGluLysLysArgAspGluLysLeuLeuProGlyLeuThrThrPro 1699

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGAT---------------------------- 819
              |||::::::            |||
Db       1700 ProLysGluValAsnAspThrPheGluAspPheLeuSerGlyLeuLysProLeuGluAla 1719

Qy        820 ------------------TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCT 861
                           :::   :::   :::   |||   :::   :::||||||
Db       1720 LeuProProLeuProAlaLeuGluProLeuGluValLysGluAspSerLysLysGluSer 1739

Qy        862 ATTACA---ACAAATGTTGAAGGA-----------------------CGAAGGGATATA 894
              ::::::      |||   |||                        |||   |||
Db       1740 ValSerLeuIleAsnPheGluSerProLeuGlnGluThrAlaSerIleProArgLysIle 1759

Qy        895 CAT---------AAAGGACATCTTGAAGAAAAG-----------AAAGATGGTTCAATA 933
                      :::|||      |||:::                        ||||||:::
Db       1760 SerProProProArgGlyPheThrGluGlnAsnLeuIleLeuAlaLeuCysGlySerLeu 1779

Qy        934 AAACCAGAACAAAAAGAAGATAAA-----------TCTGCTGACATACAAAATCAT--- 978
                 |||   ::::::::::::               |||::::::|||
Db       1780 HisTyrGluAsnGluGlnGluArgIleArgGluThrGluAlaGluValGlnProGluVal 1799

Qy        979 --------------------------ACATTAGAGACAGTA------------------ 993
                                       ||||||:::|||:::
Db       1800 ValThrThrProGluProValTyrIleThrLeuAspThrLeuAlaLeuGlnLysSerThr 1819

Qy        994 ------------AATATTTCTGATGTTAATGATTTTCAAATA------------------ 1023
                         |||   |||         :::      :::
Db       1820 SerThrSerAsnAsnSerSerGlySerGluGluIleValValMetGluGluProValLys 1839

Qy       1024 --------------AGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                           |||         ::::::      :::
Db       1840 LysLeuSerLysLysHisLysArgLeuLysGlnLeuGlnLeuGlnGlnGlnGlnGlnGln 1859

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::   |||   |||||||||:::|||                |||       |||
Db       1860 ArgGluArgGluProAspAspGluGluLeuArgProLeuIleSerIleSerMetCysAsp 1879

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
              :::      |||       :::|||   :::
Db       1880 SerGlnLeuAspValProAspIleProVal---------------------------- 1889

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                       :::   |||   |||||||||:::|||:::      |||       |||
Db       1890 ---ArgAlaMetProAspAspAspGluGlyValGluGluGluProGluThrGluProGlu 1908

Qy       1249 GAAAAAATA---------------------------------------------------- 1257
              :::   :::
Db       1909 GlnSerLeuProGluAspLeuLeuHisPheGlySerSerGlyValAlaThrAsnAlaThr 1928

Qy       1258 ---------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                      :::|||                |||   |||:::|||
Db       1929 ThrThrAspSerGluGlyProValProAlaThrThrSerAspAspAsnValValTyrGly 1948

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                 |||      :::            |||   |||      |||      :::|||
Db       1949 SerSerSerThrSerAsnAlaAlaProHisLysLeuLysThrLysLysLeuGluHisLys 1968
```

RESULT 27
AAW03627
ID AAW03627 standard; peptide; 360 AA.
XX
AC AAW03627;
XX
DT 25-MAR-2003 (revised)
DT 17-DEC-1996 (first entry)
XX
DE Human follicle stimulating hormone GPR N-terminal sequence.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomegalovirus; serotonergic.
XX
OS Homo sapiens.
XX
PN US5508384-A.
XX
PD 16-APR-1996.
XX
PF 09-SEP-1993; 93US-00118270.
XX

```
PR   10-SEP-1992;   92US-00943236.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Schuster DI,  Murphy RB;
XX
DR   WPI; 1996-208785/21.
XX
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
PT   treating schizophrenia.
XX
PS   Disclosure; Fig 8B(2); 184pp; English.
XX
CC   Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins. The peptides AAW03578-
CC   W03651 represent the N-terminal fragments of the above proteins. The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays. The polypeptide fragments retain biological activity such
CC   as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
CC   AAW02747-W02999 for examples of polypeptide fragments). The polypeptide
CC   fragments can be used in compositions for treating subjects suffering
CC   from a pathology related to a GPR abnormality e.g. a psychotic disorder
CC   such as schizophrenia. (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 360 AA;

Alignment Scores:
Pred. No.:              3.21e-16      Length:          360
Score:                  328.50        Matches:         70
Percent Similarity:     50.13%        Conservative:    118
Best Local Similarity:  18.67%        Mismatches:      130
Query Match:            13.21%        Indels:          57
DB:                     2             Gaps:            4

US-09-837-344-45 (1-1482) x AAW03627 (1-360)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||::::::  :::   |||:::|||         :::|||:::   :::|||:::
Db         33 GlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 52

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::   :::   |||::::::         :::|||:::   :::|||::::::   :::
Db         53 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 72

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                |||:::|||         :::|||:::   :::|||::::::   :::   |||:::
Db         73 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 92

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||         :::|||:::   :::|||::::::   :::   |||:::|||
Db         93 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 112

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                 :::|||:::   :::|||::::::   :::   |||::::::         :::|||
Db        113 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 132

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::   :::|||::::::   :::   |||:::|||         :::|||:::   :::
Db        133 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 152

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||::::::   :::   |||:::|||         :::|||:::   :::|||::::::
Db        153 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 172

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::   |||:::|||         :::|||:::   :::|||::::::   :::
Db        173 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 192

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::|||         :::|||:::   :::|||::::::   :::   |||:::|||

Db        193 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 212

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                      :::|||:::   :::|||::::::   :::   |||:::|||
Db        213 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 232

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::|||:::   :::|||::::::   :::   |||:::   :::   :::   ::::::
Db        233 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 252

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                    |||   ::::::|||
Db        253 GluGluGluGluGluGluGlu---------------------------------------- 259

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Db        259 ---------------------------------------------------------- 259

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
                          ::::::                    :::   :::|||
Db        260 -----------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 275

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||:::          |||
Db        276 GluGluGluGluGluGluGlu---------------------------------------- 282

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                    ||||||:::::::::         :::   |||:::::::|||::::::
Db        283 ---------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 299

Qy        961 GCTGACATACAAAATCATACATTA---GAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
                 |||                  :::      ::::::::::   |||:::
Db        300 ValAspTyrMetThrGlnAlaArgGlyGlnArgSerSerLeuAlaGluAspAsnGluSer 319

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                 |||:::      |||            |||:::|||      |||   :::|||
Db        320 SerTyrSerArgGlyPheAspMetThrTyrThrGluPheAspTyrAspLeuCysAsnGlu 339

Qy       1078 GAAGAAGAT---------GATGAAGACTTAGACGAATTTAAGCCT 1113
                    |||                 |||   |||   |||
Db        340 ValValAspValThrCysSerProLysProAspAlaPheAsnPro 354

RESULT 28
ABU53267
ID   ABU53267 standard; protein; 345 AA.
XX
AC   ABU53267;
XX
DT   14-APR-2003  (first entry)
XX
DE   Human testes-derived DKFZphtes3_50n23 homologue #1.
XX
KW   Human; gene therapy; vaccine; disease treatment; detection.
XX
OS   Homo sapiens.
XX
PN   WO200112659-A2.
XX
PD   22-FEB-2001.
XX
PF   18-AUG-2000; 2000WO-IB001496.
XX
PR   18-AUG-1999;   99US-0149499P.
PR   28-SEP-1999;   99US-0156503P.
XX
PA   (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI   Wiemann S;
XX
DR   WPI; 2001-327840/34.
XX
PT   Nucleic acids having the sequences of clones isolated from libraries of
PT   different human tissues, useful in recombinant DNA methodologies.
```

```
XX
PS   Example III; Page 902-903; 1095pp; English.
XX
CC   This invention describes novel polynucleotides and polypeptides isolated
CC   from human cDNA libraries which can be used for gene therapy or in
CC   vaccines. The polynucleotides of the invention and antibodies encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate polypeptide expression. The products of the
CC   invention may also be used to identify modulators of expression and
CC   activity and to down regulate expression and activity. The antibodies of
CC   the invention may also be used as diagnostic agents for detecting the
CC   presence of polypeptides in samples. This sequence represents a homologue
CC   of a polypeptide described in the disclosure of the invention
XX
SQ   Sequence 345 AA;

Alignment Scores:
Pred. No.:                5.83e-16      Length:          345
Score:                    325.00        Matches:         98
Percent Similarity:       50.89%        Conservative:    103
Best Local Similarity:    24.81%        Mismatches:      124
Query Match:              13.07%        Indels:          70
DB:                       4             Gaps:            14

US-09-837-344-45 (1-1482) x ABU53267 (1-345)

Qy          7 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 66
              |||      |||   ||||||   |||      :::::::::|||:::|||:::::::::
Db          2 GlnTyrArgAspLysGluGln---ArgLeuGlnArgGlnGluLeuGluGluArgArgAla 20

Qy         67 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 126
                 |||::::::::|||   :::   |||   ::::::   :::      :::|||
Db         21 ------GluGluGluGlnLeuArgArgArgLysGlyArgAspAlaGluGluPheIleGlu 38

Qy        127 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA---------GAA 177
              :::|||:::|||   :::   :::   ||||||   :::   :::|||          |||
Db         39 GluGluGlnLeuArgArgArgGluGlnGlnGluLeuLysArgGluLeuArgGluGluGlu 58

Qy        178 CAAGAGAGACGTGCTAAAGAAAAG--------------TTGCAAGAACAACAAAGCGAT 222
              |||:::||||||   :::   :::               |||||||||::::::
Db         59 GlnGlnArgArgGluArgArgGluGlnHisGluArgAlaLeuGlnGluGluGluGluGln 78

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ::::::||||||   :::|||:::   :::||||||||           |||
Db         79 LeuLeuArgGlnArgArgTrpArgGluGluProArgGluGlnGln-----------Gln 94

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
                 |||         |||::::::::::|||::::::      ||||||||||||   |||
Db         95 LeuArgArgGluLeuGluGluIleArgGluArgGluGlnArgLeuGluGlnGluGluArg 114

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
                 :::::::::|||:::   :::|||      ||||||||||||   |||   ::::::
Db        115 ArgGluGlnGlnLeuArgArgGluGlnArg---LeuGluGlnGluGluArgArgGluGln 133

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::|||:::   :::                                |||::::::::::
Db        134 GlnLeuArgArgGlu-----------------------------LeuGluGluIleArg 143

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||::::::      ||||||||||||   |||   :::::::::|||::::::::::::
Db        144 GluArgGluGlnArgLeuGluGlnGluGluArgArgGluGlnArgLeuGluGlnGluGlu 163

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
                    :::|||   :::|||      |||:::::::::|||::::::      |||
Db        164 ArgArgGluGlnGlnLeuLysArgGluLeuGluGluIleArgGluArgGluGlnArgLeu 183

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
              |||||||||   |||   ::::::   |||   |||:::   |||
Db        184 GluGlnGluGluArgArgGluGlnLeuLeuAlaGluGluValArg--------------- 198

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
                                    |||:::   :::|||   |||:::   |||
Db        199 ----------------------GluGlnAlaArgGluArgGlyGluSerLeuThrArg 210

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
                          :::||||||           |||      :::      :::
Db        211 ArgTrpGlnArgGlnLeuGlu---------SerGluAlaGlyAlaArgGlnSerLysVal 227

Qy        763 TATTATATACCA---------CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT 813
              |||      |||            :::   ||||||   ||||||
Db        228 TyrSerArgProArgArgGlnGluGluGlnSerLeuArgGlnAspGlnGluArgArgGln 247

Qy        814 AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
              |||      :::|||      :::   |||:::      |||:::
Db        248 ArgGlnGluArgGluArgGluLeuGluGluGlnAlaArgArgGlnGln---GlnTrpGln 266

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAG---------AAAGAT 924
                |||         :::      :::      |||      :::         ::::::
Db        267 AlaGluGluGluSerGluArgArgArgGlnArgLeuSerAlaArgProSerLeuArgGlu 286

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                  ::::::   |||:::::::::::::::
Db        287 ArgGlnLeuArgAlaGluGluArgGlnGluGlnGlu---------------------- 298

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                                |||      :::   |||:::|||
Db        299 --------------------------------GlnArgPheArgGluGluGluGluGln 307

Qy       1045 AGTGCTGAATATGACGATTCATTA------ATAGATGAAGAAGAAGATGATGAAGACTTA 1098
                    |||         |||      ::::::|||||||||      :::
Db        308 ArgArgGluArgArgGlnGluLeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArg 327

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
              :::      :::   :::   :::   |||:::||||||::::::
Db        328 GluArgAlaGlnGlnLeuGlnGluGluAspSerPheGlnGluAsp 342

RESULT 29
AAY29039
ID   AAY29039 standard; protein; 611 AA.
XX
AC   AAY29039;
XX
DT   24-SEP-1999  (first entry)
XX
DE   T. gondii immunogenic protein.
XX
KW   Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
KW   T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
KW   Toxoplasma oocyst.
XX
OS   Toxoplasma gondii.
XX
PN   WO9932633-A1.
XX
PD   01-JUL-1999.
XX
PF   18-DEC-1998;   98WO-US027137.
XX
PR   19-DEC-1997;   97US-00994825.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Milhausen MJ,  Lutz SB,  Ng RK;
XX
DR   WPI; 1999-418930/35.
DR   N-PSDB; AAX91242.
XX
PT   New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT   infection caused by this microorganism.
XX
PS   Claim 29; Page 227-229; 381pp; English.
XX
CC   The invention provides isolated Toxoplasma gondii nucleic acids that
CC   encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC   immunogenic proteins and antibodies to the proteins can be used to
CC   inhibit T. gondii oocyst shedding in a cat due to infection with T.
```

```
CC   gondii. They can be used for preventing T. gondii infection and for
CC   preventing the spread of T. gondii infection. They can also be used for
CC   detecting T. gondii infection. The detection method can be used to detect
CC   parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC   such as Cryptosporidium oocysts and Toxoplasma oocysts
XX
SQ   Sequence 611 AA;

Alignment Scores:
Pred. No.:              6.24e-16        Length:         611
Score:                  325.00          Matches:        84
Percent Similarity:     54.84%          Conservative:   103
Best Local Similarity:  24.63%          Mismatches:     100
Query Match:            13.07%          Indels:         54
DB:                     2               Gaps:           11

US-09-837-344-45 (1-1482) x AAY29039 (1-611)

Qy         22 GAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||::|||||||||               ||||||:::   ::::::::::::   :::
Db        292 GluArgGluArgArgArgAlaGluGluGluLysGluArgGluArgGlnGluGluGluGlu 311

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                ||:::   |||:::   :::||||||   :::|||:::|||   :::   |||:::
Db        312 ArgGluArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArg 331

Qy        130 GAG------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||      |||:::   :::||||||   ::::::::::::   |||      :::
Db        332 GluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluAspGluArgGluArg 351

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 240
              ||||||   :::|||:::         |||:::   :::   |||:::|||||||||
Db        352 ArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArg 371

Qy        241 ---------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                     |||   :::   :::|||::::::   :::   |||:::   |||:::
Db        372 ValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArgVal 391

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::||||||   ::::::::::::   :::   |||:::|||||||||      |||
Db        392 GluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgArgValGlu 411

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT--------------- 396
              :::   :::   :::|||   :::   |||:::||||||
Db        412 GluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGlu 431

Qy        397 ---AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
                  |||   :::   :::|||::::::   :::   |||      |||:::   :::|||
Db        432 GluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGluGlu 451

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
              |||   ::::::::::::   :::       :::   ||||||   :::|||:::   :::
Db        452 LysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGluGluGluLysGlu 471

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAA 567
              ::::::   :::   |||:::|||||||||         :::||||||   :::|||
Db        472 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGlu 491

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAA 615
              :::|||   :::   |||:::|||||||||                |||||:::   :::
Db        492 ArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGluArg 511

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::::::::         |||:::|||:::   :::   :::|||      :::|||:::
Db        512 GlnGluGluGluLysArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGln 531

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
              :::|||   |||
Db        532 GluGluGluGly---------------------------------------------- 535

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
                       |||   :::||||||               :::               :::
Db        536 ---------ArgGluArgGlnArgGly---------GluGluArgGluGluArgGluArg 549

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAAT 852
              :::   :::         |||:::   |||         :::   :::::::::   :::
Db        550 GluPheGlnGlnArgGluArgGluLeuLysThrArgLeuValGluLeuGlnArgGluHis 569

Qy        853 AGAGAATCTATTACAACA------AATGTTGAAGGACGAAGGGATATACAT--------- 897
                 ||||||:::   |||            :::|||   |||:::   |||
Db        570 AlaGluSerValGluThrTrpMetLysGluGlnGlyGluArgGluArgHisLeuThrGln 589

Qy        898 ------------AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                          |||   |||       |||::::::      ::::::      ::::::
Db        590 AspTrpGluArgLysLeuHisAlaPheGluGluGlnSerArgThrValLeuLeuGlnGlu 609

Qy        946 AAA 948
              :::
Db        610 Arg 610

RESULT 30
AAU25510
ID   AAU25510 standard; protein; 611 AA.
XX
AC   AAU25510;
XX
DT   17-DEC-2001  (first entry)
XX
DE   T. gondii immunogenic protein POC-23.
XX
KW   Immunogenic protein; oocyst; faeces; enteric apicomplexa oocyst;
KW   Cryptosporidium oocyst; Toxoplasma oocyst; Giardia cyst; vaccine;
KW   oocyte shedding.
XX
OS   Toxoplasma gondii.
XX
PN   US2001014447-A1.
XX
PD   16-AUG-2001.
XX
PF   18-DEC-1998;   98US-00216393.
XX
PR   19-DEC-1997;   97US-00994825.
XX
PA   (MILH/) MILHAUSEN M J.
XX
PI   Milhausen MJ;
XX
DR   WPI; 2001-529100/58.
DR   N-PSDB; AAS42565.
XX
PT   Detecting parasite oocysts or cysts in feces, comprises eluting DNA from
PT   sample into aqueous solution by heating, amplifying DNA with primers
PT   specific for oocysts or cysts being detected, and detecting amplification
PT   product.
XX
PS   Example 2; Page 91-93; 188pp; English.
XX
CC   The invention relates to detection of parasite oocysts or cysts in a
CC   faeces sample comprising contacting the sample with a solid support,
CC   drying and then washing the sample with an aqueous wash solution, adding
CC   an aqueous elution solution and eluting DNA from the sample by heating
CC   and amplifying by PCR oocyst/cyst-specific DNA and detecting the
CC   amplification products. The method is useful for detecting parasite
CC   oocysts e.g., enteric apicomplexa oocysts such as Cryptosporidium oocysts
CC   or Toxoplasma oocysts, or for detecting parasite cysts e.g. Giardia
CC   cysts. The method is also useful for developing vaccines to prevent
CC   oocyte shedding in cats. The present sequence represents an immunogenic
CC   protein from Toxoplasma gondii
XX
SQ   Sequence 611 AA;

Alignment Scores:
Pred. No.:              6.24e-16        Length:         611
Score:                  325.00          Matches:        84
Percent Similarity:     54.84%          Conservative:   103
```

```
Best Local Similarity:  24.63%          Mismatches:    100
Query Match:            13.07%          Indels:         54
DB:                     4               Gaps:           11

US-09-837-344-45 (1-1482) x AAU25510 (1-611)

Qy         22 GAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||:::|||||||||               ||||||:::   ::::::::::::   :::
Db        292 GluArgGluArgArgArgAlaGluGluGluLysGluArgGluArgGlnGluGluGluGlu 311

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                 |||:::   |||:::   :::||||||   :::|||:::|||   :::   |||:::
Db        312 ArgGluArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArg 331

Qy        130 GAG------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||      |||:::   :::||||||   ::::::::::::   |||      :::
Db        332 GluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluAspGluArgGluArg 351

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 240
              ||||||   :::|||:::          |||:::   :::   |||:::|||||||||
Db        352 ArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArg 371

Qy        241 ---------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                     |||   :::   :::|||::::::   :::   |||:::   |||:::
Db        372 ValGluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgArgVal 391

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::||||||   ::::::::::::   :::   |||:::|||||||||      |||
Db        392 GluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgArgValGlu 411

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-------------- 396
              :::   :::   :::|||   :::   |||:::||||||
Db        412 GluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGlu 431

Qy        397 ---AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
                  |||   :::   :::|||::::::   :::   |||      |||:::   :::|||
Db        432 GluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGluGlu 451

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
              |||   ::::::::::::   :::      :::   ||||||   :::|||:::   :::
Db        452 LysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGluGluGluLysGlu 471

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAA 567
                 ::::::   :::   |||:::|||||||||          :::||||||   :::|||
Db        472 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGlu 491

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAA 615
              :::|||   :::   |||:::|||||||||                ||||||:::   :::
Db        492 ArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGluArg 511

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAATTTAGAAAGAAAA 675
              :::::::::          |||:::|||:::   :::   :::|||      :::|||:::
Db        512 GlnGluGluGluLysArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGln 531

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
              :::|||   |||
Db        532 GluGluGluGly---------------------------------------------- 535

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
                     |||   :::||||||               :::               :::
Db        536 ---------ArgGluArgGlnArgGly---------GluGluArgGluGluArgGluArg 549

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAAT 852
              :::   :::          |||:::   |||          :::   :::::::::   :::
Db        550 GluPheGlnGlnArgGluArgGluLeuLysThrArgLeuValGluLeuGlnArgGluHis 569

Qy        853 AGAGAATCTATTACAACA------AATGTTGAAGGACGAAGGGATATACAT--------- 897
                 ||||||:::   |||           :::|||   |||:::   |||
Db        570 AlaGluSerValGluThrTrpMetLysGluGlnGlyGluArgGluArgHisLeuThrGln 589

Qy        898 ------------AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                         |||   |||      |||::::::      ::::::      ::::::
Db        590 AspTrpGluArgLysLeuHisAlaPheGluGluGlnSerArgThrValLeuLeuGlnGlu 609

Qy        946 AAA 948
                 :::
Db        610 Arg 610

Search completed: December 21, 2005, 10:46:06
Job time : 316.224 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 51.3343 Seconds
(without alignments)
4773.622 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093212_22815/app_query.fasta_1.3740
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_161_@runat_21122005_093212_22815 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1546	62.2	316	2	US-08-098-327E-31	Sequence 31, Appl
2	1546	62.2	316	2	US-08-462-625-31	Sequence 31, Appl
3	600	24.1	117	2	US-08-098-327E-20	Sequence 20, Appl
4	600	24.1	117	2	US-08-462-625-20	Sequence 20, Appl
5	551	22.2	107	2	US-08-098-327E-19	Sequence 19, Appl
6	551	22.2	107	2	US-08-462-625-19	Sequence 19, Appl
7	376	15.1	1898	1	US-08-056-200-94	Sequence 94, Appl
8	376	15.1	1898	1	US-08-800-644-94	Sequence 94, Appl
9	376	15.1	1898	2	US-09-538-092-1280	Sequence 1280, Ap
10	359	14.4	1162	1	US-08-728-323A-2	Sequence 2, Appli
11	359	14.4	1162	2	US-09-298-568-2	Sequence 2, Appli
12	359	14.4	1162	2	US-09-410-399-2	Sequence 2, Appli
13	359	14.4	1162	2	US-09-894-273-2	Sequence 2, Appli
14	336	13.5	608	2	US-09-270-767-32937	Sequence 32937, A
15	336	13.5	608	2	US-09-270-767-48154	Sequence 48154, A
16	325	13.1	611	2	US-09-216-393B-81	Sequence 81, Appl
17	317.5	12.8	1937	2	US-09-538-092-918	Sequence 918, App
18	317	12.7	1972	2	US-08-875-435B-3	Sequence 3, Appli
19	316	12.7	1180	2	US-09-543-681A-6436	Sequence 6436, Ap
20	316	12.7	1976	2	US-09-538-092-1078	Sequence 1078, Ap
21	315.5	12.7	1564	2	US-10-144-198-2	Sequence 2, Appli
22	315.5	12.7	1564	2	US-10-144-198-4	Sequence 4, Appli
23	315	12.7	1960	2	US-09-538-092-1077	Sequence 1077, Ap
24	315	12.7	1960	2	US-09-949-016-10872	Sequence 10872, A
25	309	12.4	1972	2	US-08-875-435B-4	Sequence 4, Appli
26	305.5	12.3	1939	2	US-09-310-187A-1	Sequence 1, Appli
27	305.5	12.3	1939	2	US-09-538-092-917	Sequence 917, App
28	303.5	12.2	924	2	US-09-248-796A-18798	Sequence 18798, A
29	303.5	12.2	1942	2	US-09-949-016-8135	Sequence 8135, Ap
30	303	12.2	1939	2	US-09-538-092-915	Sequence 915, App
31	303	12.2	1939	2	US-09-949-016-11104	Sequence 11104, A
32	303	12.2	1959	2	US-09-949-016-8134	Sequence 8134, Ap
33	302.5	12.2	1935	2	US-09-538-092-916	Sequence 916, App
34	301.5	12.1	1939	2	US-09-949-016-6925	Sequence 6925, Ap
35	301.5	12.1	1944	2	US-09-949-016-10929	Sequence 10929, A
36	301	12.1	1857	2	US-09-917-254-91	Sequence 91, Appl
37	301	12.1	1972	2	US-09-538-092-1084	Sequence 1084, Ap
38	301	12.1	1984	2	US-09-949-016-7111	Sequence 7111, Ap
39	301	12.1	1984	2	US-09-949-016-7112	Sequence 7112, Ap
40	301	12.1	1984	2	US-09-949-016-7113	Sequence 7113, Ap
41	299.5	12.0	1886	2	US-08-938-105-3	Sequence 3, Appli
42	298	12.0	735	2	US-10-164-595-80	Sequence 80, Appl
43	298	12.0	784	2	US-10-164-595-79	Sequence 79, Appl
44	298	12.0	843	2	US-10-164-595-54	Sequence 54, Appl
45	298	12.0	1938	2	US-09-949-016-6417	Sequence 6417, Ap

ALIGNMENTS

RESULT 1
US-08-098-327E-31
; Sequence 31, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; APPLICANT: DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/098,327E
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620

```
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-31

Alignment Scores:
Pred. No.:              9.15e-118      Length:         316
Score:                  1546.00        Matches:        313
Percent Similarity:     99.37%         Conservative:   1
Best Local Similarity:  99.05%         Mismatches:     2
Query Match:            62.16%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-45 (1-1482) x US-08-098-327E-31 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 2
US-08-462-625-31
; Sequence 31, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-31

Alignment Scores:
Pred. No.:              9.15e-118      Length:         316
Score:                  1546.00        Matches:        313
Percent Similarity:     99.37%         Conservative:   1
Best Local Similarity:  99.05%         Mismatches:     2
Query Match:            62.16%         Indels:         0
DB:                     2              Gaps:           0
```

```
US-09-837-344-45 (1-1482) x US-08-462-625-31 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 3
US-08-098-327E-20
; Sequence 20, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-20

Alignment Scores:
Pred. No.:                5.31e-41        Length:          117
Score:                    600.00          Matches:         116
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    99.15%          Mismatches:      0
Query Match:              24.13%          Indels:          0
DB:                       2               Gaps:            0

US-09-837-344-45 (1-1482) x US-08-098-327E-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100
```

```
Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 4
US-08-462-625-20
; Sequence 20, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  117 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-20

Alignment Scores:
Pred. No.:              5.31e-41        Length:         117
Score:                  600.00          Matches:        116
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.15%          Mismatches:     0
Query Match:            24.13%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x US-08-462-625-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 5
US-08-098-327E-19
; Sequence 19, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-19

Alignment Scores:
Pred. No.:              5.01e-37        Length:         107
Score:                  551.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            22.16%          Indels:         0
DB:                     2               Gaps:           0
```

```
US-09-837-344-45 (1-1482) x US-08-098-327E-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 6
US-08-462-625-19
; Sequence 19, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  107 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-19

Alignment Scores:
Pred. No.:                5.01e-37      Length:          107
Score:                    551.00        Matches:         106
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    99.07%        Mismatches:      0
Query Match:              22.16%        Indels:          0
DB:                       2             Gaps:            0

US-09-837-344-45 (1-1482) x US-08-462-625-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 7
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/056,200
;      FILING DATE:  30-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  94:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-056-200-94

Alignment Scores:
Pred. No.:              1.19e-22      Length:          1898
Score:                  376.00        Matches:         137
Percent Similarity:     47.39%        Conservative:    117
Best Local Similarity:  25.56%        Mismatches:      174
Query Match:            15.12%        Indels:          108
DB:                     1             Gaps:            24

US-09-837-344-45 (1-1482) x US-08-056-200-94 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::          |||:::::|||||   :::             |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                  ||||||   |||||||||   :::         :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC------------------ 219
              ||||||||||||   :::         :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                 |||:::|||||||||   :::         :::|||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::            ||||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG------------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                 :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                 ::::::||||||   :::|||:::   :::|||:::                :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 ------------------------------------GATTTAGAACAAGAG--------- 540
                                                 ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||      |||:::      |||:::                     :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                 |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
                 ||||||::::::   |||||||||:::       |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT-------------- 717
              |||   |||   |||            :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||   |||            :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::             |||      |||                 ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                        :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||      :::         |||:::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::          :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                       :::                 ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                  |||          |||:::   |||||||||::::::          :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                       :::      :::   :::|||            |||:::
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::       |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::          :::       :::|||            |||       ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 8
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
```

; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-644-94

Alignment Scores:

Pred. No.:	1.19e-22	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	1	Gaps:	24

US-09-837-344-45 (1-1482) x US-08-800-644-94 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||        |||||||||  |||  ::::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::||||||   ::            |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              ||||||   |||||||||  :::        :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------------------- 219
              ||||||||||||   :::        :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
              |||:::|||||||||   :::        :::||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::                ||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG----------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                  |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                :::::|||||||   :::|||:::   :::|||:::                :::
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluAlaArg 713

Qy        526 ----------------------------------GATTTAGAACAAGAG--------- 540
                                                 ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||       |||:::      |||:::                  :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
                |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
              ||||||::::::   |||||||||:::      |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
              |||   |||   |||            :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||   |||          :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::            |||   |||                  ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                  :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                    :::|||   :::   |||      :::          |||::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                    :::   :::   :::|||::::::          :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                       :::                  ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                 |||          |||:::   |||||||||::::::          :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                                  :::      :::   :::|||              |||:::
Db        935 ------------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||          :::      |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::          :::      :::|||              |||      ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   ::::::::::::   :::|||:::|||:::   ::::::
```

```
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 9
US-09-538-092-1280
; Sequence 1280, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1280
;   LENGTH: 1898
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number Q07283
US-09-538-092-1280

Alignment Scores:
Pred. No.:              1.19e-22        Length:         1898
Score:                  376.00          Matches:        137
Percent Similarity:     47.39%          Conservative:   117
Best Local Similarity:  25.56%          Mismatches:     174
Query Match:            15.12%          Indels:         108
DB:                     2               Gaps:           24

US-09-837-344-45 (1-1482) x US-09-538-092-1280 (1-1898)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||||||||          |||||||||   |||   :::::::::|||:::   |||
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::||||||   :::             |||||||||
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
                ||||||   |||||||||   :::         :::|||::::::      ||||||
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
              ||||||||||||   :::         :::||||||:::
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                 |||:::|||||||||   :::         :::||||||:::      ||||||
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGlnGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
              |||   ||||||   :::|||:::::::::               ||||||||||||
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG-----------------CAAGAACAACAAAGCGATTTA--- 378
              |||      |||:::|||                :::|||::::::   :::|||
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                      |||||||||   |||      |||:::|||:::   :::|||   :::
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA--------------CAAAGC 474
                  ::::::||||||   :::|||:::   :::|||:::               :::

Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
              :::   |||:::|||||||||   :::|||      :::|||:::|||
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 -----------------------------------GATTTAGAACAAGAG--------- 540
                                                  ||||||   |||
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
              |||::::::|||       |||:::      |||:::                     :::
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA------------------------ 615
                 |||::::::||||||   :::   :::||||||
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
              ||||||::::::   |||||||||:::      |||   :::          |||
Db        774 ArgGlnGlnGlnGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
              |||   |||   |||             :::   |||       |||
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
                      ::::::   |||   |||                  :::|||      :::
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
              :::|||   :::             |||      |||                  ||||||
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
              :::      :::|||:::      :::                        :::   :::
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                  :::|||   :::   |||      :::         |||:::::::::|||
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                  :::   :::   :::|||::::::            :::   |||||||||
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                  :::                  ::::::::::::
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
                 |||          |||:::   |||||||||::::::         :::|||
Db        917 ArgLysGlu------GlnGlnLeuLeuGlnGluGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                              :::      :::   :::|||             |||:::
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
              |||   |||   |||   ::::::   |||   |||       |||   |||       |||
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||         :::       |||   :::   :::|||   ||||||   ::::::
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
              ::::::            :::      :::|||                |||      ||||||
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   :::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022
```

```
RESULT 10
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma-Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1162 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-728-323A-2

Alignment Scores:
Pred. No.:                2.69e-21        Length:          1162
Score:                    359.00          Matches:         94
Percent Similarity:       51.19%          Conservative:    121
Best Local Similarity:    22.38%          Mismatches:      117
Query Match:              14.44%          Indels:          88
DB:                       1               Gaps:            12

US-09-837-344-45 (1-1482) x US-08-728-323A-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||         :::::::::   |||       |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||           :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
                :::       ||||||||||||   :::   |||||||:::       ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::             :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                  ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
                :::   :::     :::|||           |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                |||   |||
Db        828 ----------------------------GlnGluGlnGlu---------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                        :::   :::|||      :::
Db        832 -----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                |||:::         |||                 |||       ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                :::     |||||||||   ::::::         :::   |||:::::::|::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                ::::::   :::   |||::::::         |||   ::::::         |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                             :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu---------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
                ::::::|||                 :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                          :::|||:::|||:::      :::|||      ||||||   |||
Db        923 ---------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 11
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
```

```
;  APPLICANT: Ballestas, Mary E.
;  APPLICANT: Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/298,568
;  CURRENT FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,422
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

Alignment Scores:
Pred. No.:              2.69e-21        Length:         1162
Score:                  359.00          Matches:        94
Percent Similarity:     51.19%          Conservative:   121
Best Local Similarity:  22.38%          Mismatches:     117
Query Match:            14.44%          Indels:         88
DB:                     2               Gaps:           12

US-09-837-344-45 (1-1482) x US-09-298-568-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::|||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||          :::::::::   |||       |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::      |||:::|||           :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                ||:::::::       |||:::   |||::::::|||          ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::       ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::              :::       ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::          :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      :::      :::|||            |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu-------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                  |||   |||
Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                  :::   :::|||      :::
Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
                 |||:::          |||                  |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
                 :::      |||||||||   ::::::         :::   |||::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
                 ::::::   :::   |||::::::         |||   ::::::          |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
                 ||:::::::::::::::|||:::|||                          :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
                 ::::::|||               :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln------------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                           :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

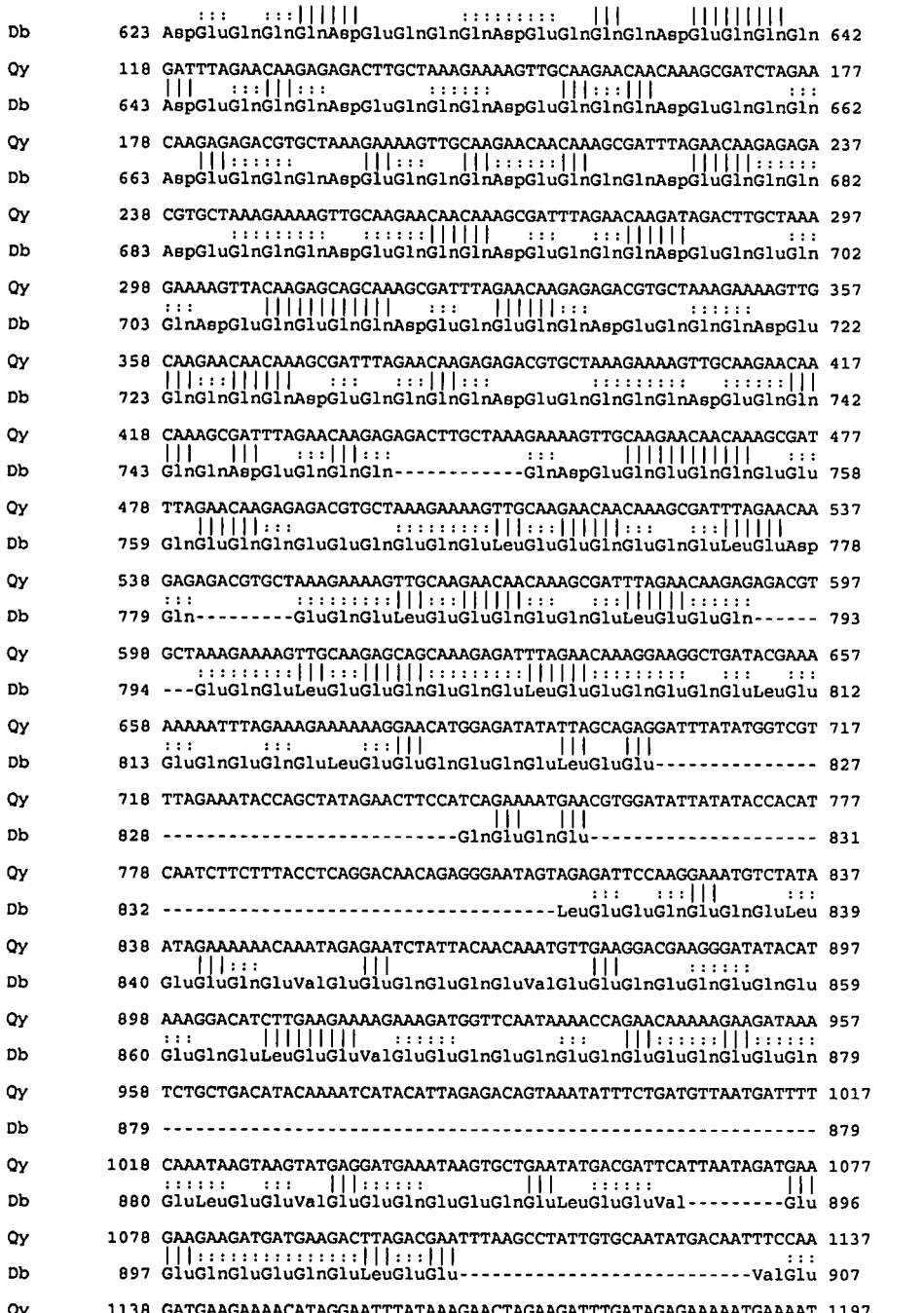
RESULT 12
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
;  APPLICANT:  Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION:  Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE:  UM-03778
;  CURRENT APPLICATION NUMBER: US/09/410,399
;  CURRENT FILING DATE:  1999-10-01
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

Alignment Scores:
Pred. No.:              2.69e-21        Length:         1162
Score:                  359.00          Matches:        94
Percent Similarity:     51.19%          Conservative:   121
Best Local Similarity:  22.38%          Mismatches:     117
Query Match:            14.44%          Indels:         88
DB:                     2               Gaps:           12

US-09-837-344-45 (1-1482) x US-09-410-399-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::|||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
```



```
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177

Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237

Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297

Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357

Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417

Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477

Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537

Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597

Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657

Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717

Db        813 GluGluGlnGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837

Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897

Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957

Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGluGlnGluGluGlnGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077

Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137

Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197

Db        908 GluGlnGluGln------------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257

Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 13
US-09-894-273-2
; Sequence 2, Application US/09894273
; Patent No. 6756203
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT:  Ballestas, Mary E.
;  APPLICANT:  Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION:  VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/894,273
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/109,422
;  PRIOR FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

Alignment Scores:
Pred. No.:              2.69e-21      Length:          1162
Score:                  359.00        Matches:           94
Percent Similarity:     51.19%        Conservative:     121
Best Local Similarity:  22.38%        Mismatches:       117
Query Match:            14.44%        Indels:            88
DB:                     2             Gaps:              12

US-09-837-344-45 (1-1482) x US-09-894-273-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63

Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117

Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177

Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237

Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297

Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357

Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417

Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477

Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
```

```
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597

Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657

Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717

Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777

Db        828 ----------------------------GlnGluGlnGlu-------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837

Db        832 -----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897

Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957

Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ------------------------------------------------------------ 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077

Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137

Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197

Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257

Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937
```

RESULT 14
US-09-270-767-32937
; Sequence 32937, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32937
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32937

Alignment Scores:

Pred. No.:	1.82e-19	Length:	608
Score:	336.00	Matches:	70
Percent Similarity:	64.94%	Conservative:	80
Best Local Similarity:	30.30%	Mismatches:	73

Query Match:	13.51%	Indels:	8
DB:	2	Gaps:	4

US-09-837-344-45 (1-1482) x US-09-270-767-32937 (1-608)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db        374 LysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArgLysGluGlu 393

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Db        394 GluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGluGluArgArg 413

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Db        414 LysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArgLysGluGlu 433

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Db        434 GluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGluGluArgLys 452

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Db        453 AlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArgArgGluGlu 472

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Db        473 LysArgArgArgGluGlu---------GluGluArgArgLysGluGluGluArgArgGlu 489

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTGCAAGAACAA 417

Db        490 GluGluGluLysArgLysGluGluGluArgArgLysAspGluGluArgArgArgGluGlu 509

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477

Db        510 GluLysArgLysGluGluGluArgArgGluLysGluArgArgArgGluGluGlyLysArg 529

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537

Db        530 LysGluGluGluArgArgGluLysGluArgArgArgGluGluGluLysArgLysGluGlu 549

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597

Db        550 GluArgArgGluLysGluArgArgAspGluGluArgArgArgGluGluGluArgArgArg 569

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA---------AGGAAGGCT 648

Db        570 GluGluGluArgArgArgGluGluGluArgArgArgGluGluGluArgArgGluGlu 589

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681

Db        590 GluArgArgArgGluGluGluArgLysArgGlu 600
```

RESULT 15
US-09-270-767-48154
; Sequence 48154, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48154
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48154

Alignment Scores:

Pred. No.:	1.82e-19	Length:	608

Score:	336.00	Matches:	70
Percent Similarity:	64.94%	Conservative:	80
Best Local Similarity:	30.30%	Mismatches:	73
Query Match:	13.51%	Indels:	8
DB:	2	Gaps:	4

US-09-837-344-45 (1-1482) x US-09-270-767-48154 (1-608)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::         |||:::|||||||||   :::|||:::   :::|||:::
Db        374 LysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArgLysGluGlu 393

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              :::         |||::::::|||      :::|||:::   :::|||::::::
Db        394 GluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGluGluArgArg 413

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||:::   |||     :::::::::   :::|||::::::         |||:::
Db        414 LysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArgLysGluGlu 433

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||:::   :::|||:::   :::|||   :::         |||:::||||||:::
Db        434 GluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGluGluArgLys 452

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||:::|||:::   ::::::::::::         |||:::   |||     :::|||
Db        453 AlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArgArgGluGlu 472

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              |||   :::   ::::::         |||:::   |||:::   :::|||:::   :::
Db        473 LysArgArgArgGluGlu---------GluGluArgArgLysGluGluGluArgArgGlu 489

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTGCAAGAACAA 417
              |||::::::         |||:::|||||||||       :::|||:::   :::|||:::
Db        490 GluGluGluLysArgLysGluGluGluArgArgLysAspGluGluArgArgArgGluGlu 509

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              :::         |||:::||||||       ||||||:::   :::|||:::
Db        510 GluLysArgLysGluGluGluArgArgGluLysGluArgArgArgGluGluGlyLysArg 529

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
              |||:::|||||||||   ||||||:::   :::|||::::::         |||:::
Db        530 LysGluGluGluArgArgGluLysGluArgArgArgGluGluGluLysArgLysGluGlu 549

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              |||||||||   ||||||:::       |||::::::         |||:::   ||||||
Db        550 GluArgArgGluLysGluArgArgAspGluGluArgArgArgGluGluGluArgArgArg 569

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA---------AGGAAGGCT 648
              :::|||:::   :::|||:::::|||      |||:::         |||:::
Db        570 GluGluGluArgArgArgGluGluGluArgArgArgGluGluGluArgArgArgGluGlu 589

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::   ::::::      |||||||||:::|||
Db        590 GluArgArgArgGluGluGluArgLysArgGlu 600
```

RESULT 16
US-09-216-393B-81
; Sequence 81, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393B
; CURRENT FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393B-81

Pred. No.:	1.42e-18	Length:	611
Score:	325.00	Matches:	84
Percent Similarity:	54.84%	Conservative:	103
Best Local Similarity:	24.63%	Mismatches:	100
Query Match:	13.07%	Indels:	54
DB:	2	Gaps:	11

US-09-837-344-45 (1-1482) x US-09-216-393B-81 (1-611)

```
Qy         22 GAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||:::|||||||||                 ||||||:::   ::::::::::::   :::
Db        292 GluArgGluArgArgArgAlaGluGluGluLysGluArgGluArgGlnGluGluGluGlu 311

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
              |||:::   |||:::   :::||||||   :::|||:::|||   :::   |||:::
Db        312 ArgGluArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArg 331

Qy        130 GAG------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||      |||:::   :::||||||   ::::::::::::   |||         :::
Db        332 GluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluAspGluArgGluArg 351

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 240
              ||||||   :::|||:::          |||:::   :::   |||:::|||||||||
Db        352 ArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArg 371

Qy        241 ---------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                        |||   :::   :::|||::::::   :::   |||:::   |||:::
Db        372 ValGluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgArgVal 391

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
              :::||||||   ::::::::::::   :::   |||:::|||||||||      |||
Db        392 GluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArgArgArgValGlu 411

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT--------------- 396
              :::   :::   :::|||   :::   |||:::||||||
Db        412 GluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGlu 431

Qy        397 ---AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
                 |||   :::   :::|||::::::   :::   |||       |||:::   :::|||
Db        432 GluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGluGlu 451

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
              |||   ::::::::::::   :::       :::   ||||||   :::|||:::   :::
Db        452 LysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGluGluGluLysGlu 471

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAA 567
              ::::::   :::   |||:::|||||||||         :::|||||   :::|||
Db        472 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGlu 491

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAA 615
              :::|||   :::   |||:::|||||||||               ||||||:::   :::
Db        492 ArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGluArg 511

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::::::::         |||:::|||:::   :::   :::|||      :::|||:::
Db        512 GlnGluGluGluLysArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGln 531

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
              :::|||   |||
Db        532 GluGluGluGly---------------------------------------------- 535

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
                       |||   :::||||||               :::           :::
Db        536 ---------ArgGluArgGlnArgGly---------GluGluArgGluGluArgGluArg 549

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAAT 852
              :::   :::         |||:::   |||       :::   :::::::::   :::
Db        550 GluPheGlnGlnArgGluArgGluLeuLysThrArgLeuValGluLeuGlnArgGluHis 569
```

```
Qy        853 AGAGAATCTATTACAACA------AATGTTGAAGGACGAAGGGATATACAT--------- 897
Db        570 AlaGluSerValGluThrTrpMetLysGluGlnGlyGluArgGluArgHisLeuThrGln 589

Qy        898 ------------AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
Db        590 AspTrpGluArgLysLeuHisAlaPheGluGluGlnSerArgThrValLeuLeuGlnGlu 609

Qy        946 AAA 948
Db        610 Arg 610
```

RESULT 17
US-09-538-092-918
; Sequence 918, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 918
; LENGTH: 1937
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P13535
US-09-538-092-918

Alignment Scores:

Pred. No.:	6.74e-18	Length:	1937
Score:	317.50	Matches:	156
Percent Similarity:	42.08%	Conservative:	131
Best Local Similarity:	22.87%	Mismatches:	184
Query Match:	12.77%	Indels:	211
DB:	2	Gaps:	35

US-09-837-344-45 (1-1482) x US-09-538-092-918 (1-1937)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540
Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db       1250 LysMetCysArgSerLeuGlu------------AspGlnValSerGluLeuLysThrLys 1265

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
Db       1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG------TCTATA 837
Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy        933 ------------------------------------------------------------ 933
Db       1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy        934 --------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
Db       1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
```

```
                    |||   |||         :::                 :::   |||   :::
Db       1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT---------- 1086
                :::      |||   :::|||   ::::::|||   |||
Db       1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy       1087 ----------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                          |||:::   |||   |||:::|||
Db       1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA---------------------- 1152
                :::|||:::      |||   |||
Db       1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy       1153 ----------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176
                                     :::|||:::|||           ||||||
Db       1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy       1177 TTGATAGAGAAAAATGAAAATTTA-----------GATGATTTAGATGAAGGAATAGAA 1224
              |||      :::|||:::||||||              ||||||   |||   |||
Db       1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy       1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
              :::      :::::::::   |||   ||||||||||||||      :::   ||||||
Db       1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy       1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
                :::          :::      :::   ||||||      |||      |||
Db       1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy       1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
                :::         ::::::|||:::   |||   :::      :::
Db       1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
                |||   :::   |||   ||||||            |||:::|||      :::
Db       1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy       1459 ACTAAA 1464
                |||:::
Db       1594 ThrArg 1595

RESULT 18
US-08-875-435B-3
; Sequence 3, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
;  APPLICANT: Hasegawa, Kazuhide
;  APPLICANT:  Arakawa, Emi
;  APPLICANT:  Oda, Shoji
;  APPLICANT:  Matsuda, Yuzuru
;  APPLICANT:  Takahashi, Katsuhito
;  APPLICANT:  Sugahara, Michihiro
;  APPLICANT:  Ishiyama, Haruo
;  TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
;  TITLE OF INVENTION:  FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
;  TITLE OF INVENTION:  VECTOR DNA, MICROORGANSIM CARRYING THE RECOMBINANT DNA, AND
;  TITLE OF INVENTION:  AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
;  TITLE OF INVENTION:  RECOMBINANT DNA
;  FILE REFERENCE: 07898-013001
;  CURRENT APPLICATION NUMBER: US/08/875,435B
;  CURRENT FILING DATE:  1997-07-25
;  PRIOR APPLICATION NUMBER: PCT/JP96/00134
;  PRIOR FILING DATE: 1996-01-25
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-08-875-435B-3
```

```
Alignment Scores:
Pred. No.:              7.42e-18        Length:         1972
Score:                  317.00          Matches:        144
Percent Similarity:     42.02%          Conservative:   114
Best Local Similarity:  23.45%          Mismatches:     180
Query Match:            12.75%          Indels:         176
DB:                     2               Gaps:           27

US-09-837-344-45 (1-1482) x US-08-875-435B-3 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                            ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-----------------------------------------------AGCGAT 120
              ::::::                                               |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA------------------------ 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                       :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ------------------------------------AGACTTGCTAAA------GAAAAG 303
                                                 :::|||||||||      |||:::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG--------------------------- 336
              ||||||          :::   |||::::::|||
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
                    |||      :::      :::   :::   |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   :::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::            :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                         |||   |||:::              |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||  :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                  ||||||                             |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251
```

```
Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::      ::::::      :::||||||
Db       1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu---------------------- 1264

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                              ||||||
Db       1265 ----------------------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                   |||      :::      |||   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||:::|||                  ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ::::::      ||||||   :::          |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              :::|||      |||   :::   :::|||:::|||               :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||||||||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||      ::::::|||:::          |||         |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 ::::::           |||:::|||:::|||               |||   :::
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::      :::   ::::::|||      |||   ||||||         ||||||
Db       1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------- 1290
              |||            |||   ||||||||||||:::
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1291 ------------------------------------AATAATTTTAAACCAAATGAT 1311
                                                 :::|||   :::
Db       1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT---------------------- 1350
              |||      ::|||:::   :::   :::   ||||||
Db       1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy       1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
                    |||:::      :::   :::||||||      |||:::|||      :::
Db       1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy       1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
                        |||   |||:::      :::||||||
Db       1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
```

```
RESULT 19
US-09-543-681A-6436
; Sequence 6436, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6436
;   LENGTH: 1180
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-6436

Alignment Scores:
Pred. No.:              8.35e-18      Length:          1180
Score:                  316.00        Matches:         98
Percent Similarity:     56.02%        Conservative:    51
Best Local Similarity:  36.84%        Mismatches:      77
Query Match:            12.71%        Indels:          40
DB:                     2             Gaps:            14

US-09-837-344-45 (1-1482) x US-09-543-681A-6436 (1-1180)

Qy          7 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTGCAAGAA 57
              :::|||          ||||||||||||   |||:::         |||:::   :::
Db        592 LysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeuGlnAlaGluGlnGluLysAla 611

Qy         58 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAAAAGTTACAAGAG 108
              :::|||          ||||||:::|||||||||:::         |||:::   :::
Db        612 LysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeuGlnAlaGluGlnGluLysAla 631

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              :::|||          |||||||||||||||||       :::||||||   :::|||
Db        632 LysGlnGluLysAlaGluGlnGluArgLeuAla------ArgLeuGlnAlaGluGlnGlu 649

Qy        169 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAG-----------------------TTG 204
                    :::||||||:::      :::||||||                        |||
Db        650 LysAlaLysGlnGluLysAlaGluGlnGluLysAlaGluGlnGluArgLeuAlaArgLeu 669

Qy        205 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG--------- 255
              |||   :::|||          :::||||||:::      :::|||:::|||
Db        670 GlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAlaArgLeu 689

Qy        256 ---------------CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                           :::   :::|||         ||||||:::|||||||||
Db        690 GlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAla------ 707

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::||||||   :::|||          :::||||||:::      :::|||:::|||
Db        708 ArgLeuGlnAlaGluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluArgLeuAla 727

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
                    |||:::      |||||||||:::      :::||||||   ::::::::::::
Db        728 ArgLeuGlnAla------GluGlnGluLysAlaLysGlnGluLysAlaGluGlnGluLys 745

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              :::      ||||||||||||||||||      |||||||||   :::|||
Db        746 Ala------GluGlnGluArgLeuAla------LysLeuGlnAlaGluGlnGluLysAla 761

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||   |||      :::||||||   :::|||          |||||||||
Db        762 GluGlnGluArgLeuAla------ArgLeuGlnAlaGluGlnGluLysThrGluGlnGlu 779

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||   |||      :::||||||   :::|||          ||||||||||||   |||
Db        780 ArgLeuAla------ArgLeuGlnAlaGluGlnGluLysAlaGluGlnGluArgLeuAla 797

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
                    :::||||||   :::|||          ||||||   :::          :::
Db        798 ------ArgLeuGlnAlaGluGlnGluLysThrGluGlnGluArgLeu---AlaArgLeu 814

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                    |||:::::|||          :::   ||||||      |||
Db        815 GlnAlaGluGlnGluLysAlaGluGlnGluArgLeuAlaArg---LeuGlnAlaAspLys 833

Qy        721 GAAATACCAGCTATAGAA 738
              |||:::|||   :::|||
Db        834 GluLeuProProValGlu 839
```

```
RESULT 20
US-09-538-092-1078
; Sequence 1078, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1078
;   LENGTH: 1976
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Alignment Scores:
Pred. No.:                8.94e-18        Length:          1976
Score:                    316.00          Matches:         144
Percent Similarity:       40.48%          Conservative:    109
Best Local Similarity:    23.04%          Mismatches:      180
Query Match:              12.71%          Indels:          192
DB:                       2               Gaps:            25

US-09-837-344-45 (1-1482) x US-09-538-092-1078 (1-1976)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||   |||   |||   |||   |||:::   ::::::   |||   ::::::|||:::
Db       1122 GluLeuGlnGluAspPheGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLys 1141

Qy         64 AGCGATTTAGAACAAGATAGACTTGCT---AAAGAAAAGTTACAAGAG----------- 108
                ||||||   ::::::      |||   |||   :::|||::::::
Db       1142 ArgAspLeuSerGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThr 1161

Qy        109 ------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 162
                   |||   :::|||      :::|||      ::::::::::   |||
Db       1162 ThrAlaAlaGlnGlnGluLeuArgThrLysArg------GluGlnGluValAlaGluLeu 1179

Qy        163 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 198
              :::      ||||||:::|||   :::      |||
Db       1180 LysLysAlaLeuGluGluGluThrLysAsnHisGluAlaGlnIleGlnAspMetArgGln 1199

Qy        199 ------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 246
                         |||:::|||            ||||||||   :::|||   |||
Db       1200 ArgHisAlaThrAlaLeuGluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLys 1219

Qy        247 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT------AGACTTGCTAAAGAA 300
                |||::::::   :::      ||||||   |||         ||||||   |||
Db       1220 AlaAsnLeuGluLysAsnLysGlnGlyLeuGluThrAspAsnLysGluLeuAlaCysGlu 1239

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA--------- 351
                  ||||||          |||   |||   :::   |||:::
Db       1240 ValLysValLeuGlnGlnValLysAlaGluSerGluHisLysArgLysLysLeuAspAla 1259

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
              ::::::||||||      :::   :::   :::      |||   :::   :::|||
Db       1260 GlnValGlnGluLeuHisAlaLysValSerGluGlyAspArgLeuArgValGluLeuAla 1279

Qy        412 GAACAACAAAGCGATTTAGAACAAGAG------------------------------ 438
              |||:::   |||   |||:::   |||
Db       1280 GluLysAlaSerLysLeuGlnAsnGluLeuAspAsnValSerThrLeuLeuGluGluAla 1299

Qy        439 ---------------AGACTTGCTAAAGAA-----------------AAGTTGCAAGAA 465
                            :::   ||||||:::                :::||||||:::
Db       1300 GluLysLysGlyIleLysPheAlaLysAspAlaAlaSerLeuGluSerGlnLeuGlnAsp 1319

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
                |||      |||::::::|||   |||   |||      |||      :::
Db       1320 ThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnLeuSerSerArgIleArg 1339

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                ||||||:::|||         |||      |||||||||||||||         |||
Db       1340 GlnLeuGluGluGlu---------LysAsnSerLeuGlnGluGlnGln---------Glu 1353

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
              :::|||      |||::::::   |||::::::|||         |||:::   :::
Db       1354 GluGluGluGluAlaArgLysAsnLeuGluLysGlnValLeuAlaLeuGlnSerGlnLeu 1373

Qy        646 GCTGATACGAAAAAAAATTTAGAA---------------------------------- 669
              ||||||||||||||   ::::::
Db       1374 AlaAspThrLysLysLysValAspAspAspLeuGlyThrIleGluSerLeuGluGluAla 1393

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              :::||||||         |||      ||||||   |||      |||||||||
Db       1394 LysLysLysLeuLeuLysAsp------AlaGluAlaLeuSerGlnArgLeuGlu------ 1409

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                                :::   :::|||
Db       1410 ---------------------------------------------GluLysAlaLeu 1413

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA------------ 837
                 |||            :::::::::      :::      :::
Db       1414 AlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuThr 1433

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA------------ 885
              ::::::      ::::::      :::   :::|||:::|||   :::
Db       1434 ValAspLeuAspHisGlnArgGlnValAlaSerAsnLeuGluLysLysGlnLysLysPhe 1453

Qy        886 ---------------------AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
                                    :::   |||          :::   ||||||:::
Db       1454 AspGlnLeuLeuAlaGluGluLysSerIleSerAlaArgTyrAlaGluGluArgAspArg 1473

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA--------------- 969
                  :::      :::||||||   |||:::      :::
Db       1474 AlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGlu 1493

Qy        970 --------------------------------CAAAATCATACATTAGAGACAGTAAAT 996
                                               ||||||      |||
Db       1494 GluAlaLeuGluAlaLysGluGluPheGluArgGlnAsnLysGlnLeuArg--------- 1510

Qy        997 ATTTCTGATGTTAATGATTTT------------------------------------- 1017
                 :::|||:::   |||
Db       1511 ---AlaAspMetGluAspLeuMetSerSerLysAspAspValGlyLysAsnValHisGlu 1529

Qy       1018 ------------------------------------------------CAAATAAGT 1026
                                                             |||:::
Db       1530 LeuGluLysSerLysArgAlaLeuGluGlnGlnValGluGluMetArgThrGlnLeuGlu 1549

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
              :::   |||||||||:::   |||      |||      |||      |||
Db       1550 GluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluValAsnMet 1569

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                          :::|||:::      :::|||      :::   |||:::||||||
Db       1570 GlnAlaMetLysAlaGlnPheGluArgAspLeuGlnThrArgAspGluGlnAsnGluGlu 1589

Qy       1147 AAC------------ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
                            |||         :::||||||||||      :::|||
Db       1590 LysLysArgLeuLeuIleLysGlnValArgGluLeuGluAlaGluLeuGluAspGluArg 1609

Qy       1195 ------------------------------------------AATTTAGATGATTTA 1209
                                                        :::|||   ||||||
Db       1610 LysGlnArgAlaLeuAlaValAlaSerLysLysLysMetGluIleAspLeuLysAspLeu 1629
```

```
Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
Db       1630 GluAlaGlnIleGlu---AlaAlaAsnLysAlaArgAspGluValIleLysGlnLeuArg 1648

Qy       1270 AAATATGAA---AAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
Db       1649 LysLeuGlnAlaGlnMetLysAsp----------------------------TyrGln 1658

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAG-----------GTTAATAAGGAAAAG 1374
Db       1659 ArgGluLeuGluGluAlaArgAlaSerArgAspGluIlePheAlaGlnSerLysGluSer 1678

Qy       1375 GAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATC 1434
Db       1679 GluLysLysLeuLysSerLeu-----------------GluAlaGluIleLeuGlnLeu 1692

Qy       1435 GTGGATGAGTTATCT 1449
Db       1693 GlnGluGluLeuAla 1697

RESULT 21
US-10-144-198-2
; Sequence 2, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies Inc
;  TITLE OF INVENTION: Regulated Prostate Cance Genes
;  FILE REFERENCE: 9U 105 R1
;  CURRENT APPLICATION NUMBER: US/10/144,198
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1564
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-198-2

Alignment Scores:
Pred. No.:              9.52e-18        Length:         1564
Score:                  315.50          Matches:        115
Percent Similarity:     43.69%          Conservative:   110
Best Local Similarity:  22.33%          Mismatches:     163
Query Match:            12.69%          Indels:         127
DB:                     2               Gaps:           20

US-09-837-344-45 (1-1482) x US-10-144-198-2 (1-1564)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        629 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 648

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114
Db        649 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 668

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        669 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 687

Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db        688 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 707

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
Db        708 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 727

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
Db        728 AspArg--------------------AspArgAspHisAspArgGluArgGluArgGlu 740

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402

Db        741 ---ArgGluArgAspArgGluLysGluArgGluArgGluArgGluGluArgGluArgGlu 759

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
Db        760 ArgGluArgGluArgGluArgGluArgGluArgGluArg---------------------- 772

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAA 519
Db        773 -----------------GluArgGluArgAlaArgGluArgAspLysGluArgGluArg 786

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
Db        787 GlnArgAspTrpGluAspLysAspLysGlyArgAspAspArgArgGluLysArgGluGlu 806

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db        807 IleArgGluAspArgAsnProArgAspGlyHisAspGluArgLysSerLysLysArgTyr 826

Qy        640 AGGAAGGCTGATACG---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
Db        827 ArgAsnGluGlySerProSerProArgGlnSerProLysArgArgArgGluHisSer--- 845

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
Db        845 ------------------------------------------------------------ 845

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
Db        846 ------------------------------------ProAspSerAspAlaTyrAsn 852

Qy        811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
Db        853 SerGlyAspAspLysAsn-----------GluLysHisArgLeuLeuSerGlnValVal 868

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--- 927
Db        869 ArgProGlnGluSerArgSerLeuSerProSerHisLeuThrGluAspArgGlnGlyArg 888

Qy        928 ------------TCAATAAAACCAGAACAAAAAGAAGAT---------------------- 954
Db        889 TrpLysGluGluAspArgLysProGluArgLysGluSerSerArgArgTyrGluGluGln 908

Qy        955 -----------------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
Db        909 GluLeuLysGluLysValSerSerValAspLysGlnArgGluGlnThrGluIleLeuGlu 928

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
Db        929 SerSerArgMetArgAlaGlnAspIleIleGlyHisHisGlnSerGluAspArgGluThr 948

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
Db        949 SerAspArgAlaHisAspGluAsnLysLysLysAlaLysIleGlnLys---LysProIle 967

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
Db        968 LysLys------------LysLysGluAspAspValGlyIleGluArg------------ 979

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
Db        980 ----------------------------------GlyAsnIleGluThrThrSerGlu 987

Qy       1237 GAA---------TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
Db        988 AspGlyGlnValPheSerProLysLysGlyGlnLysLysLysSerIleGluLysLysArg 1007

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
Db       1008 LysLysSer------LysGlyAspSerAspIleSerAspGluGluAlaAlaGlnGlnSer 1025

Qy       1348 AATGATAAGCAG--------------------GTTAATAAGGAAAAGGAAAAATTCATA 1386
```

```
Db       1026 LysLysLysArgGlyProArgThrProProIleThrThrLysGluGlu---------- 1041

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                 |||      :::   :::|||   |||   ||||||:::
Db       1042 ------LeuValGluMetCysAsnGlyLysAsnGlyIleLeuGlu 1054

RESULT 22
US-10-144-198-4
; Sequence 4, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies Inc
;  TITLE OF INVENTION: Regulated Prostate Cance Genes
;  FILE REFERENCE: 9U 105 R1
;  CURRENT APPLICATION NUMBER: US/10/144,198
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 1564
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-198-4

Alignment Scores:
Pred. No.:              9.52e-18      Length:          1564
Score:                  315.50        Matches:         115
Percent Similarity:     43.69%        Conservative:    110
Best Local Similarity:  22.33%        Mismatches:      163
Query Match:            12.69%        Indels:          127
DB:                     2             Gaps:            20

US-09-837-344-45 (1-1482) x US-10-144-198-4 (1-1564)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::::::::::::   :::   |||:::::::::|||          ::::::   |||
Db        629 ArgAspArgArgAspAsnArgGluArgAspGlnArgProSerSerProIleArgHisGln 648

Qy         61 ------CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAA 114
                   :::||||||:::|||         :::|||:::      ::::::
Db        649 GlyArgAsnAspGluLeuGluArgAspGluArgArgGluGluArgArgValAspArgVal 668

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
                |||          ||||||   |||:::|||:::   :::|||::::::   |||
Db        669 AspAspArgArgAspGluArg---AlaArgGluArgAspArgGluArgGluArgAspArg 687

Qy        175 GAACAAGAGAGA---------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
              |||:::||||||         |||   :::::::::   :::|||::::::   :::|||
Db        688 GluArgGluArgGluArgGluArgGluArgAspArgGluArgGluLysGluArgGluLeu 707

Qy        226 GAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||:::||||||   |||   :::|||:::   :::|||::::::   |||   |||:::
Db        708 GluArgGluArgAlaArgGluArgGluArgGluArgGluLysGluArgAspArgGluArg 727

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
              ||||||                 ::::::      |||   |||:::||||||
Db        728 AspArg--------------------AspArgAspHisAspArgGluArgGluArgGlu 740

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 402
                 :::|||:::   :::|||::::::   :::   |||:::|||   |||   :::|||
Db        741 ---ArgGluArgAspArgGluLysGluArgGluArgGluArgGluGluArgGluArgGlu 759

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::   :::|||::::::   :::   |||:::||||||
Db        760 ArgGluArgGluArgGluArgGluArgGluArgGluArg--------------------- 772

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAA 519
                              |||:::||||||   |||   :::::::|||   :::|||:::
Db        773 -----------------GluArgGluArgAlaArgGluArgAspLysGluArgGluArg 786

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||   |||   |||   :::   :::   ::::::      :::|||::::::   :::
Db        787 GlnArgAspTrpGluAspLysAspLysGlyArgAspAspArgArgGluLysArgGluGlu 806

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              :::   ::::::|||      ::::::         |||::::::
Db        807 IleArgGluAspArgAsnProArgAspGlyHisAspGluArgLysSerLysLysArgTyr 826

Qy        640 AGGAAGGCTGATACG---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
              |||         :::         :::::::::   :::|||::::::||||||
Db        827 ArgAsnGluGlySerProSerProArgGlnSerProLysArgArgArgGluHisSer--- 845

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db        845 ------------------------------------------------------------ 845

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
                                                        |||      :::      |||
Db        846 ----------------------------------------ProAspSerAspAlaTyrAsn 852

Qy        811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
              |||   |||   |||                 ||||||            |||
Db        853 SerGlyAspAspLysAsn-----------GluLysHisArgLeuLeuSerGlnValVal 868

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT--- 927
                  :::      |||   :::         ||||||   |||   :::   |||
Db        869 ArgProGlnGluSerArgSerLeuSerProSerHisLeuThrGluAspArgGlnGlyArg 888

Qy        928 ------------TCAATAAAACCAGAACAAAAAGAAGAT---------------------- 954
                          |||||||||:::||||||
Db        889 TrpLysGluGluAspArgLysProGluArgLysGluSerSerArgArgTyrGluGluGln 908

Qy        955 ------------------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT 996
                               |||   |||   |||              |||   :::
Db        909 GluLeuLysGluLysValSerSerValAspLysGlnArgGluGlnThrGluIleLeuGlu 928

Qy        997 ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
                 |||   :::              |||      :::                |||
Db        929 SerSerArgMetArgAlaGlnAspIleIleGlyHisHisGlnSerGluAspArgGluThr 948

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
                |||          ||||||   :::             :::   :::   |||||||||
Db        949 SerAspArgAlaHisAspGluAsnLysLysLysAlaLysIleGlnLys---LysProIle 967

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
                 :::             :::   |||:::::::::||||||   :::
Db        968 LysLys------------LysLysGluAspAspValGlyIleGluArg------------ 979

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
                                                          ||||||   ::::||||||
Db        980 ----------------------------------GlyAsnIleGluThrThrSerGlu 987

Qy       1237 GAA---------TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                 :::             |||   :::|||   :::|||   |||      ||||||   :::
Db        988 AspGlyGlnValPheSerProLysLysGlyGlnLysLysLysSerIleGluLysLysArg 1007

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
                    :::             |||      :::   ||||||         :::
Db       1008 LysLysSer------LysGlyAspSerAspIleSerAspGluGluAlaAlaGlnGlnSer 1025

Qy       1348 AATGATAAGCAG--------------------GTTAATAAGGAAAAGGAAAAATTCATA 1386
                          |||:::                        ||||||:::
Db       1026 LysLysLysArgGlyProArgThrProProIleThrThrLysGluGlu---------- 1041

Qy       1387 AAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                 |||      :::   :::|||   |||   ||||||:::
Db       1042 ------LeuValGluMetCysAsnGlyLysAsnGlyIleLeuGlu 1054

RESULT 23
US-09-538-092-1077
; Sequence 1077, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
```

```
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1077
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Alignment Scores:
Pred. No.:                1.08e-17      Length:          1960
Score:                    315.00        Matches:         145
Percent Similarity:       44.63%        Conservative:    100
Best Local Similarity:    26.41%        Mismatches:      168
Query Match:              12.67%        Indels:          136
DB:                       2             Gaps:            28

US-09-837-344-45 (1-1482) x US-09-538-092-1077 (1-1960)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||:::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::        |||            ||||||   :::|||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
                              |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::|||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                 |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                ::::||||||      |||:::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||       :::   :::|||:::||||||         :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                                |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||       |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                    ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||          |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                              ||||||||||||||      :::
Db       1221 --------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
              ::::::      :::            :::|||:::|||         |||
Db       1231 AsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATGTCTATAATAGAA---------------------------------AAAACAAAT 852
              |||      :::   |||                                 :::|||
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                       |||   ::::::         |||         |||   |||
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG-----------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
                  :::                      ||||||   |||            |||
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              :::|||            :::|||            :::   ::::::::::|||
Db       1311 GlnAspThrGlnGluLeuGluGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                    ::::::   :::   |||||||||   ::::::   :::   :::   |||
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::|||||||||         :::|||:::      |||   |||
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                          :::          :::   |||::::::||||||   :::
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
                 |||      :::::::::          ::::::|||      |||   ||||||
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                    |||||||||:::           |||   ||||||||||||
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      |||   |||   ::|||      :::   ::::::
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||:::::::::||||||
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446

RESULT 24
US-09-949-016-10872
; Sequence 10872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10872
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10872

Alignment Scores:
Pred. No.:                1.08e-17        Length:          1960
Score:                    315.00          Matches:         145
Percent Similarity:       44.63%          Conservative:    100
Best Local Similarity:    26.41%          Mismatches:      168
Query Match:              12.67%          Indels:          136
DB:                       2               Gaps:            28

US-09-837-344-45 (1-1482) x US-09-949-016-10872 (1-1960)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||:::::::||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::          |||           ||||||   :::||||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
                               |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::||||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                  |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                 :::||||||      |||::::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
                 :::|||:::   |||      :::   :::|||:::||||||          :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
                 |||:::|||                              |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
                 :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
                 |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
                 :::|||      |||:::::::::                          ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
                 |||      :::   ::::::|||          ||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                 |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                 ||||||||||||||      :::
Db       1221 -------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
                 ::::::      :::          :::|||:::|||         |||
Db       1231 AsnGluValLysValLeuGluGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATGTCTATAATAGAA--------------------------------AAAACAAAT 852
                 |||      :::  |||                                :::|||
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                       |||   ::::::          |||          |||   |||
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG------------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
                 :::                        ||||||   |||          |||
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
                 :::|||            :::|||            :::   :::::::::|||
Db       1311 GlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                       ::::::   :::   |||||||||   ::::::   :::   :::   |||
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                    :::|||||||||         :::|||:::      |||   |||
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                          :::          :::   |||::::::||||||   :::
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
                 |||      :::::::::          ::::::|||      |||   ||||||
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                    |||||||||:::          |||   ||||||||||||
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
                 :::      |||      |||   |||   :::|||      :::   ::::::
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||:::::::::||||||
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446

RESULT 25
US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
```

```
; GENERAL INFORMATION:
;  APPLICANT: Hasegawa, Kazuhide
;  APPLICANT:  Arakawa, Emi
;  APPLICANT:  Oda, Shoji
;  APPLICANT:  Matsuda, Yuzuru
;  APPLICANT:  Takahashi, Katsuhito
;  APPLICANT:  Sugahara, Michihiro
;  APPLICANT:  Ishiyama, Haruo
;  TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
;  TITLE OF INVENTION:  FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
;  TITLE OF INVENTION:  VECTOR DNA, MICROORGANSIM CARRYING THE RECOMBINANT DNA, AND
;  TITLE OF INVENTION:  AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
;  TITLE OF INVENTION:  RECOMBINANT DNA
;  FILE REFERENCE: 07898-013001
;  CURRENT APPLICATION NUMBER: US/08/875,435B
;  CURRENT FILING DATE:  1997-07-25
;  PRIOR APPLICATION NUMBER: PCT/JP96/00134
;  PRIOR FILING DATE: 1996-01-25
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Alignment Scores:
Pred. No.:                3.31e-17      Length:          1972
Score:                    309.00        Matches:         150
Percent Similarity:       41.95%        Conservative:    113
Best Local Similarity:    23.92%        Mismatches:      162
Query Match:              12.42%        Indels:          202
DB:                       2             Gaps:            30

US-09-837-344-45 (1-1482) x US-08-875-435B-4 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||         ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::     |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                   |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAAGAA--------------------- 198
                           |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        199 ------------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAA 228
                                           :::|||:::::::::::::       :::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        229 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT--- 285
              :::      :::   |||   |||:::      :::   |||||||||   :::
Db       1052 GluLeuGluLysLeuLysArgLysMetAspGlyGluAlaSerAspLeuHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
                                               :::|||||||||      |||:::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAA------------CAAGAGAGACGTGCTAAAGAA 351
              ||||||         :::   ||||||           |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuGluGluThrSerGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
                                                |||:::::::|||                                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   :::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::            :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::           |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAA 570
                      ||||||   |||   |||                |||:::|||   ||||||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnValValGluGluLeuThrGluGln 1211

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
                   ||||||||   :::|||||||||      |||   :::   ::::::
Db       1212 ---------LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThr 1228

Qy        631 TTAGAACAAAGGAAGGCTGAT-----------------------------ACGAAAAAA 660
              ||||||:::      ||||||                                |||:::
Db       1229 LeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1248

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              :::|||   ||||||:::                                   :::|||
Db       1249 GluValGluHisLysLysLys--------------------------------LysLeu 1257

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              |||:::         ||||||   |||:::         |||
Db       1258 GluValGlnLeuGlnGluLeuGlnSerLysCysSerAspGly----------------- 1271

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
                                                  |||         |||::::::
Db       1272 ------------------------------GluArgAlaArgAlaGluLeuAsn------ 1279

Qy        841 GAAAAAACAAATAGA--------------GAATCTATTACAACAAATGTTGAAGGACGA 885
              :::|||   ::::::                |||||:::|||
Db       1280 AspLysValHisLysLeuGlnAsnGluValGluSerValThr----------------- 1293

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA--------- 936
                       |||   |||   |||   :::      :::||||||
Db       1294 --------------GlyMetLeuSerGluAlaGluGlyLysAlaIleLysLeuAlaLys 1308

Qy        937 --------------CCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                           |||   :::|||                :::|||      |||
Db       1309 GluValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuGlnGluGluThr 1328

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
              :::   :::|||:::|||               ::::::   :::   |||||||||
Db       1329 ArgGlnLysLeuAsnValSerThr-----------LysLeuArgGlnLeuGluAspGlu 1344

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
                          :::||||||      :::|||:::   ||||||:::::::::
Db       1345 --------------ArgAsnSerLeu------GlnGluGlnLeuAspGluGluMetGlu 1357

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
                    :::   :::   ::::::   :::      :::   :::
Db       1358 AlaLysGlnAsnLeuGluArgHisIleSerThrLeuAsnIleGlnLeuSerAspSerLys 1377

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221
              |||:::|||:::|||                |||:::|||::::::      :::
Db       1378 LysLysLeuGlnAspPheAlaSerThrValGluSerLeuGluGluGlyLysLysArgPhe 1397

Qy       1222 GAAAAATCATCAGAAGAATTATCT---------GAAGAAAAAATAAAAAAAGGAAAGAAA 1272
              :::|||       |||   |||:::         |||||||||                |||
Db       1398 GlnLysGluIleGluSerLeuThrGlnGlnTyrGluGluLysAlaAlaAlaTyrAspLys 1417

Qy       1273 TATGAAAAAACAAAGGAT---------------------------------------- 1290
                 ||||||||||||:::
```

```
Db       1418 LeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAsp 1437

Qy       1291 ------------------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332
                              :::|||   :::         |||      :::|||:::
Db       1438 AsnGlnArgGlnLeuValSerAsnLeuGluLysLysGlnLysLys---PheAspGlnLeu 1456

Qy       1333 ATTAAAAAATATAAAAAT-------------------------GATAAGCAGGTTAAT 1365
              :::   :::   ||||||                       |||:::
Db       1457 LeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArgAspArgAlaGluAla 1476

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
              :::   :::||||||      |||:::|||      :::                |||
Db       1477 GluAlaArgGluLysGluThrLysAlaLeu---SerLeuAlaArgAlaLeuGluGluAla 1495

Qy       1426 TTACAGATCGTGGATGAGTTA 1446
              |||:::      :::||||||
Db       1496 LeuGluAlaLysGluGluLeu 1502

RESULT 26
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
;  APPLICANT:  Benichou, Gilles
;  APPLICANT:  Fedoseyeva, Eugenia
;  TITLE OF INVENTION:  Involvement of Autoantigens in Cardiac
;  TITLE OF INVENTION:  Graft Rejection
;  FILE REFERENCE:  UCSF-090
;  CURRENT APPLICATION NUMBER: US/09/310,187A
;  CURRENT FILING DATE:  1999-05-12
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-310-187A-1

Alignment Scores:
Pred. No.:              6.35e-17        Length:         1939
Score:                  305.50          Matches:        148
Percent Similarity:     45.60%          Conservative:   132
Best Local Similarity:  24.10%          Mismatches:     181
Query Match:            12.28%          Indels:         153
DB:                     2               Gaps:           35

US-09-837-344-45 (1-1482) x US-09-310-187A-1 (1-1939)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        929 GluAspGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 948

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
                ||||:::|||::::::|||                       ::::::   |||
Db        949 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 968

Qy         94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                       |||   |||:::   :::   |||:::
Db        969 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 988

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        989 GluIleIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnAla 1008

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                  ||||||:::   |||                       ::::::|||   |||
Db       1009 LeuAspAspLeuGlnValGluGluAspLysValAsnSerLeuSerLysSerLysValLys 1028

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGACGTGCTAAAGAA 249
              |||:::::|||      |||            ||||||||||:::::   :::
Db       1029 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysValArgMet 1048

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
                 |||:::      :::      ||||||  |||   :::|||   :::|||   :::
Db       1049 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuThrGlnGluSerIle 1068

Qy        307 CAAGAGCAGCAAAGCGAT---TTAGAACAAGAGAGACGTGCTAAAGAAAAG--------- 354
              :::   :::::::|||   |||:::   |||   :::   |||:::|||
Db       1069 MetAspLeuGluAsnAspLysLeuGlnLeuGluGluLysLeuLysLysLysGluPheAsp 1088

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA------------CGTGCTAAAGAA 402
              :::   :::|||   |||   :::|||   |||:::            :::   ::::::
Db       1089 IleAsnGlnGlnAsnSerLysIleGluAspGluGlnAlaLeuAlaLeuGlnLeuGlnLys 1108

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              ||||||:::|||   |||:::                                 :::::::::
Db       1109 LysLeuLysGluAsnGlnAla-----------------------------ArgIleGlu 1118

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||   :::   :::||||||   ||||||   |||:::   |||:::::::::   :::
Db       1119 GluLeuGluGluGluLeuGluAlaGluArgThrAlaArgAlaLysValGluLysLeuArg 1138

Qy        523 AGCGATTTAGAACAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAA------------ 567
              |||||||||   :::|||                 |||:::|||:::|||
Db       1139 SerAspLeuSerArgGluLeuGluGluIleSerGluArgLeuGluGluAlaGlyGlyAla 1158

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                     :::|||      ::::::  :::   :::   |||:::   :::|||
Db       1159 ThrSerValGlnIleGluMetAsnLysLysArgGluAlaGluPheGlnLysMetArgArg 1178

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACAT 684
              |||||||||:::                  :::                 :::|||:::|||
Db       1179 AspLeuGluGluAlaThrLeuGlnHisGluAlaThrAlaAlaAlaLeuArgLysLysHis 1198

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA---ATACCAGCTATAGAACTT 741
                 |||   :::|||||||   |||      :::::::::   :::      ::::::
Db       1199 AlaAspSerValAlaGlu---LeuGlyGluGlnIleAspAsnLeuGlnArgValLysGln 1217

Qy        742 CCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAAC 801
                 |||   |||:::      :::                    |||      |||
Db       1218 LysLeuGluLysGluLysSerGluPhe---------------LysLeuGluLeuAsp--- 1231

Qy        802 AGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCT 861
                    :::         :::  |||      |||      |||   |||   |||
Db       1232 ------AspValThrSerAsnMetGluGlnIleIleLysAlaLysAlaAsnLeuGluLys 1249

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA------ 915
              ::::::      :::|||   :::   :::   ::::::      |||||||||
Db       1250 ValSerArgThrLeuGluAspGlnAlaAsnGluTyrArgValLysLeuGluGluAlaGln 1269

Qy        916 ----------------------------------------AAGAAAGATGGTTCAATA 933
                                                     ::::::|||   :::
Db       1270 ArgSerLeuAsnAspPheThrThrGlnArgAlaLysLeuGlnThrGluAsnGlyGluLeu 1289

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTA 993
                  |||   |||:::|||   |||   |||         |||      :::
Db       1290 ---AlaArgGlnLeuGluGluLysGluAlaLeuIleSerGlnLeuThrArgGlyLysLeu 1308

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA------------ 1041
              :::   :::         :::         :::   |||:::|||
Db       1309 SerTyrThrGlnGlnMetGluAspLeuLysArgGlnLeuGluGluGluGlyLysAlaLys 1328

Qy       1042 --------------------ATAAGTGCTGAATATGACGATTCATTAATAGATGAA--- 1077
                                   ||||||   :::|||      |||:::   |||
Db       1329 AsnAlaLeuAlaHisAlaLeuGlnSerAlaArgHisAspCysAspLeuLeuArgGluGln 1348

Qy       1078 ---GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTC 1134
                 ||||||:::   |||      |||   :::   ::::::         |||
Db       1349 TyrGluGluGluThrGluAlaLysAlaGluLeuGlnArgValLeuSerLysAlaAsnSer 1368

Qy       1135 CAA-----------------------GATGAAGAAAACATAGGAATTTATAAAGAACTA 1170
              :::                      :::   :::   |||         :::||||||
Db       1369 GluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGlnArgThrGluGluLeu 1388

Qy       1171 GAAGATTTGATAGAGAAA---AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              |||:::      :::|||      :::   |||   |||   :::|||   :::|||
```

```
Db       1389 GluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAspAlaGluGluAlaValGluAla 1408

Qy       1228 TCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAA------ 1278
              :::   :::   |||      ||||||   |||        :::   |||
Db       1409 ValAsnAlaLysCysSerSerLeuGluLysThrLysHisArgLeuGlnAsnGluIleGlu 1428

Qy       1279 ------------------------------------------------AAAACAAAGGAT 1290
                                                             |||
Db       1429 AspLeuMetValAspValGluArgSerAsnAlaAlaAlaAlaAlaLeuAspLysLysGln 1448

Qy       1291 AATAATTTTAAACCAAATGATAAAAGTTTG-----------------TATGATGAGCAT 1332
              ||||||         ||||||   |||                 |||:::|||
Db       1449 ArgAsnPhe---------AspLysIleLeuAlaGluTrpLysGlnLysTyrGluGluSer 1465

Qy       1333 ATTAAAAAATATAAAAATGATAAG----------------------CAGGTTAATAAG 1368
                  :::   ::::::   :::                     ::::::   |||
Db       1466 GlnSerGluLeuGluSerSerGlnLysGluAlaArgSerLeuSerThrGluLeuPheLys 1485

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATA------TTTGACGGAGACAATGAA 1422
              |||      :::   :::||||||   |||:::      |||      :::|||:::
Db       1486 LeuLysAsnAlaTyrGluGluSerLeuGluHisLeuGluThrPheLysArgGluAsnLys 1505

Qy       1423 ATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
              ||||||      :::|||:::|||   |||:::|||:::
Db       1506 AsnLeuGln------GluGluIleSer---AspLeuThrGlu 1516

RESULT 27
US-09-538-092-917
; Sequence 917, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 917
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P13533
US-09-538-092-917

Alignment Scores:
Pred. No.:              6.35e-17      Length:          1939
Score:                  305.50        Matches:         148
Percent Similarity:     45.60%        Conservative:    132
Best Local Similarity:  24.10%        Mismatches:      181
Query Match:            12.28%        Indels:          153
DB:                     2             Gaps:            35

US-09-837-344-45 (1-1482) x US-09-538-092-917 (1-1939)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   |||||:::::::::
Db        929 GluAspGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 948

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
                 |||:::|||::::::|||                         ::::::   |||
Db        949 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 968

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::

Db        969 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 988

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        989 GluIleIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnAla 1008

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                     :::::::|||   |||
Db       1009 LeuAspAspLeuGlnValGluGluAspLysValAsnSerLeuSerLysSerLysValLys 1028

Qy        202 TTGCAAGAACAACAAAGCGAT-----------TTAGAACAAGAGAGACGTGCTAAAGAA 249
              |||::::::|||       |||           ||||||||||||::::::   :::
Db       1029 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysValArgMet 1048

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
              |||:::      :::      ||||||  |||  :::|||  :::|||  :::
Db       1049 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuThrGlnGluSerIle 1068

Qy        307 CAAGAGCAGCAAAGCGAT---TTAGAACAAGAGAGACGTGCTAAAGAAAAG--------- 354
                 :::   ::::::|||   |||:::   |||   :::   |||:::|||
Db       1069 MetAspLeuGluAsnAspLysLeuGlnLeuGluGluLysLeuLysLysLysGluPheAsp 1088

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA-----------CGTGCTAAAGAA 402
              :::   :::|||   |||   :::|||   |||:::           :::   ::::::
Db       1089 IleAsnGlnGlnAsnSerLysIleGluAspGluGlnAlaLeuAlaLeuGlnLeuGlnLys 1108

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              ||||||:::|||   |||:::                                :::::::::
Db       1109 LysLeuLysGluAsnGlnAla----------------------------ArgIleGlu 1118

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||   :::   :::||||||   ||||||   |||:::   |||::::::::   :::
Db       1119 GluLeuGluGluGluLeuGluAlaGluArgThrAlaArgAlaLysValGluLysLeuArg 1138

Qy        523 AGCGATTTAGAACAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAA----------- 567
              |||||||||   :::|||                |||:::|||:::|||
Db       1139 SerAspLeuSerArgGluLeuGluGluIleSerGluArgLeuGluGluAlaGlyGlyAla 1158

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                       :::|||      ::::::   :::   :::   |||:::   :::|||
Db       1159 ThrSerValGlnIleGluMetAsnLysLysArgGluAlaGluPheGlnLysMetArgArg 1178

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACAT 684
              |||||||||:::                :::                :::|||:::|||
Db       1179 AspLeuGluGluAlaThrLeuGlnHisGluAlaThrAlaAlaAlaLeuArgLysLysHis 1198

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA---ATACCAGCTATAGAACTT 741
                 |||   :::||||||   |||      :::::::::   :::      ::::::
Db       1199 AlaAspSerValAlaGlu---LeuGlyGluGlnIleAspAsnLeuGlnArgValLysGln 1217

Qy        742 CCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAAC 801
                    |||   |||:::      :::                  |||      |||
Db       1218 LysLeuGluLysGluLysSerGluPhe--------------LysLeuGluLeuAsp--- 1231

Qy        802 AGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCT 861
                    :::         :::   |||      |||      |||   |||   |||
Db       1232 ------AspValThrSerAsnMetGluGlnIleIleLysAlaLysAlaAsnLeuGluLys 1249

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA------ 915
              ::::::      :::|||   :::   :::   ::::::      |||||||||
Db       1250 ValSerArgThrLeuGluAspGlnAlaAsnGluTyrArgValLysLeuGluGluAlaGln 1269

Qy        916 ----------------------------------------AAGAAAGATGGTTCAATA 933
                                                     ::::::|||   :::
Db       1270 ArgSerLeuAsnAspPheThrThrGlnArgAlaLysLeuGlnThrGluAsnGlyGluLeu 1289

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTA 993
                 |||   |||:::|||   |||   |||          |||          :::
Db       1290 ---AlaArgGlnLeuGluGluLysGluAlaLeuIleSerGlnLeuThrArgGlyLysLeu 1308

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA----------- 1041
              :::   :::          :::          :::   |||:::|||
Db       1309 SerTyrThrGlnGlnMetGluAspLeuLysArgGlnLeuGluGluGluGlyLysAlaLys 1328
```

```
Qy       1042 --------------------ATAAGTGCTGAATATGACGATTCATTAATAGATGAA--- 1077
                                   ||||||   :::|||      |||:::   |||
Db       1329 AsnAlaLeuAlaHisAlaLeuGlnSerAlaArgHisAspCysAspLeuLeuArgGluGln 1348

Qy       1078 ---GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTC 1134
                  ||||||:::   |||         |||   :::   ::::::         |||
Db       1349 TyrGluGluGluThrGluAlaLysAlaGluLeuGlnArgValLeuSerLysAlaAsnSer 1368

Qy       1135 CAA----------------------GATGAAGAAAACATAGGAATTTATAAAGAACTA 1170
                :::                     :::   :::   |||         :::||||||
Db       1369 GluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGlnArgThrGluGluLeu 1388

Qy       1171 GAAGATTTGATAGAGAAA---AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
                |||:::      :::|||      :::   |||   |||   :::|||   :::|||
Db       1389 GluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAspAlaGluGluAlaValGluAla 1408

Qy       1228 TCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAA------ 1278
                  :::   :::   |||      ||||||   |||         :::   |||
Db       1409 ValAsnAlaLysCysSerSerLeuGluLysThrLysHisArgLeuGlnAsnGluIleGlu 1428

Qy       1279 ------------------------------------------------AAAACAAAGGAT 1290
                                                              |||
Db       1429 AspLeuMetValAspValGluArgSerAsnAlaAlaAlaAlaAlaLeuAspLysLysGln 1448

Qy       1291 AATAATTTTAAACCAAATGATAAAAGTTTG-----------------TATGATGAGCAT 1332
                |||||||      ||||||   |||                 |||:::|||
Db       1449 ArgAsnPhe---------AspLysIleLeuAlaGluTrpLysGlnLysTyrGluGluSer 1465

Qy       1333 ATTAAAAAATATAAAAATGATAAG----------------------CAGGTTAATAAG 1368
                  :::   ::::::   :::                      ::::::   |||
Db       1466 GlnSerGluLeuGluSerSerGlnLysGluAlaArgSerLeuSerThrGluLeuPheLys 1485

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATA------TTTGACGGAGACAATGAA 1422
                |||      :::   :::||||||   |||:::      |||      :::|||:::
Db       1486 LeuLysAsnAlaTyrGluGluSerLeuGluHisLeuGluThrPheLysArgGluAsnLys 1505

Qy       1423 ATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
                ||||||      :::|||:::|||   |||:::|||:::
Db       1506 AsnLeuGln------GluGluIleSer---AspLeuThrGlu 1516
```

RESULT 28
US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Alignment Scores:

Pred. No.:	8.37e-17	Length:	924
Score:	303.50	Matches:	129
Percent Similarity:	43.24%	Conservative:	111
Best Local Similarity:	23.24%	Mismatches:	196
Query Match:	12.20%	Indels:	119
DB:	2	Gaps:	21

US-09-837-344-45 (1-1482) x US-09-248-796A-18798 (1-924)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||:::         :::|||      |||      :::|||:::::::::::::|||:::
Db        325 GlnGlnPheThrLysGluLeuIleAlaGluSerGluSerLysGlnGlnValGluGluGlu 344

Qy         61 CAAAGCGATTTAGAA------------CAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                  :::|||:::               :::|||               :::|||:::
Db        345 LeuGluAsnLeuLysThrIleHisAlaLysAspSerAlaArgIleLeuGluLeuGluThr 364

Qy        109 CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              |||   ||||||      :::|||:::         :::                   |||
Db        365 GlnLeuSerAspAlaAlaLysGluLysSerGluSerAspTyrLysLeuThrAspThrSer 384

Qy        169 GATCTAGAACAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAACAACAAAGCGATTTA 225
              ::::::      :::   :::::::::  |||   |||:::      :::   |||
Db        385 GluIleValAsnAspLeuLysSerGlnIleGluThrLeuLysAlaAsnLeuAsnLysLeu 404

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAA 282
              |||:::||||||      :::   |||      :::|||   |||:::|||   |||
Db        405 GluGluGluArgGluIleGlnAsnLysLysLeuAspGlnValSerGluLeuLysGluLeu 424

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
              |||      |||:::|||                     |||:::||||||
Db        425 LysGluLeuLysValGluGluLeuSerAsnAsnLeuLeuLysLeuGlnGlnGluLeuHis 444

Qy        343 GCTAAAGAA---AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 399
                ||||||   ::|||   ||||||      |||                     |||
Db        445 ArgLysGluIleGluLeuAsnGluGlnLeuGluLysLeuHisGlyValSerAlaAspLys 464

Qy        400 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 459
              |||||||||   :::|||         :::||||||   ::::::              |||
Db        465 GluLysLeuLeuGlnGln------SerIleGluGlnSerLysIleTyrIleAsnSerLeu 482

Qy        460 CAAGAACAACAAAGCGATTTAGAACAA--------------------------GAGAGA 492
              :::::::|||   :::      ::::::                        |||:::
Db        483 LysAspGlnAsnAlaLysThrAspArgAsnTrpGluIleLysTyrGlyThrLeuGluGln 502

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                  ::::::   ::::::|||      ::::::   ::::::::::::
Db        503 AsnGlnGlnLysLeuValGluGluHisLeuAlaAsnArgGlnArgAspGlnLeuGluIle 522

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAG 609
              |||||||||   :::   :::::::::||||||   :::   :::|||   |||   :::
Db        523 GluLysLeuValLysSerLysThrGluLeuGluAlaAspAsnGlnAlaLeuLysAsnGln 542

Qy        610 TTGCAAGAG---------CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||   |||         |||         |||             ::::::   ||||||
Db        543 LeuThrGluIleSerLysGlnPheSerThrLeuGlyValGluSerSerGluLeuLysLys 562

Qy        661 AATTTAGAAAGA---------AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711
                |||||      |||   :::       :::|||:::   ::::::|||
Db        563 GlnLeuGluGluThrGlnIleLysTyrAspGlnAlaGluIleValHisGlnGluLeuLys 582

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771
              ::::::|||
Db        583 AlaGlnIleGlu---------------------------------------------- 586

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG 831
                      |||   :::::::::   |||                 ||||||:::
Db        587 ------------LysLeuLeuLysGluSerAlaGlyLysAspGlyGlnLeuLysGluLeu 602

Qy        832 TCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGAT 891
              :::                     :::|||         |||         ::::::
Db        603 Asn--------------------GlnSerHisGluLysPheValAsnGluLeuLysAsn 615

Qy        892 ATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951
                |||:::      :::::|||   ||||||      ||||||      :::::::::
Db        616 GluHisGluLysGlnValLysGluThrLysAspGlnIleIleLysGluMetGluGluLys 635

Qy        952 GATAAATCTGCT--------------GACATACAAAATCATACATTAGAGACAGTAAAT 996
                :::   |||               |||      |||      ::::::   :::|||
Db        636 HisGlnGlnAlaIleLysGluIleGluAspSerHisAsnGluAsnIleLysGluIleAsn 655

Qy        997 ----------------------------ATTTCTGATGTTAATGAT------------ 1014
```

```
                                          |||      :::|||:::
Db        656 AsnGluHisGluAsnLysAlaLysCysIleIleAspSerLeuAsnGluGluIleGluGlu 675

Qy       1015 ------TTTCAAATAAGTAAGTATGAGGATGAA-----------------ATAAGTGCT 1050
                     |||:::          |||   |||               :::
Db        676 LeuThrSerGlnLeuLysAsnAlaGluSerGluLysAsnThrLeuGlnSerLeuLysLeu 695

Qy       1051 GAATATGACGATTCATTAATAGAT----------------------GAAGAAGAAGAT 1086
              ||||||::::::   :::|||                        |||:::|||
Db        696 GluTyrGluAsnGluIleIleAlaTyrLysSerLysIleAspGlnLeuGluLysGluSer 715

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                 |||:::|||   |||::::::      :::|||          |||      |||
Db        716 AlaGluAsnLeuLysGluTyrGluAlaLysLeuGlnSerMetLysPheAspLeuGluSer 735

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGAT---------------------------- 1176
                 ::::::   |||   |||:::|||
Db        736 AspLeuAlaIleGluLysGlnLeuArgLysAspAspGlyGlnAspPheGluAsnGlnIle 755

Qy       1177 ------------------TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAA--- 1215
                           :::   ||||||:::      |||   :::   ||||||
Db        756 GluLysLeuAsnGlnLeuValThrGluLysAspLeuGlnLeuSerGluLysAspGluGlu 775

Qy       1216 ------GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
                    |||:::|||     |||||||||   :::   |||   |||       :::
Db        776 IleAlaSerIleLysLysTyrMetGluGluLeuGluAspThrLysAlaLysLeuGluGlu 795

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
              |||         |||         |||      |||   ||||||      :::
Db        796 Lys------SerThrThrGlnAsnGlyMetLysLeuLeuAspAsnSerLeuThrLysLys 813

Qy       1330 CAT------ATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
              |||      :::             :::   |||      |||   :::   :::
Db        814 HisAlaAlaValValAlaGluLeuSerGlyLysIleThrAsnGlyAspAlaAsp------ 831

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
                                   ||||||||||||||   |||
Db        832 --------------------IleAspGlyAspAsnGluProLeu 839
```

RESULT 29
US-09-949-016-8135
; Sequence 8135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8135
; LENGTH: 1942
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8135

Alignment Scores:

Pred. No.:	9.24e-17	Length:	1942
Score:	303.50	Matches:	154
Percent Similarity:	41.46%	Conservative:	142
Best Local Similarity:	21.57%	Mismatches:	185
Query Match:	12.20%	Indels:	233
DB:	2	Gaps:	38

US-09-837-344-45 (1-1482) x US-09-949-016-8135 (1-1942)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        934 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 953

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                |||:::|||::::::|||                          ::::::   |||
Db        954 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 973

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        974 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 993

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        994 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1013

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                    ||||||:::   |||                    :::||||||   |||
Db       1014 LeuAspAspLeuGlnMetGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1033

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||            ||||||||||||:::
Db       1034 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuCysMet 1053

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||   ||||||                     ||||||
Db       1054 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1073

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
                 |||   |||   |||                :::|||   |||::::::   :::
Db       1074 MetAspThrGluAsnAspLysGlnGlnLeuAsnGluLysLeuLysLysLysGluPheGlu 1093

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                                         ||||||         ::::::   ::::::
Db       1094 MetSerAsnLeuGlnGlyLysIleGluAspGluGlnAlaLeuAlaIleGlnLeuGlnLys 1113

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||         :::::::::|||   :::   ::::::|||   ||||||
Db       1114 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1133

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
              ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1134 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1153

Qy        448 AAAGAAAAGTTGCAAGAA-----------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
                 |||:::|||:::|||                  :::   :::|||       ::::::
Db       1154 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1173

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
              :::   :::   |||:::   :::   |||||||||:::
Db       1174 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluSerThrLeuGlnHisGlu 1193

Qy        540 ------------------------------------------------------------ 540

Db       1194 AlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1213

Qy        541 -----------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                        :::|||   |||:::||||||:::::::::::|||:::|||:::
Db       1214 IleAspSerLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1233

Qy        589 GAG-------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                                 :::||||||         :::
Db       1234 GluIleAsnAspLeuAlaSerAsnMetGluThrValSerLysAlaLysAlaAsnPheGlu 1253

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTAGAAAGA 672
              :::      |||   ||||||   :::   ::::::   |||   |||       |||:::
Db       1254 LysMetCysArgThrLeuGluAspGlnLeuSerGluIleLysThrLysGluGluGluGln 1273

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ::::::          ::::::   |||:::           ||||||          :::
Db       1274 GlnArgLeuIleAsnGluLeuSerAlaGlnLys------AlaArgLeuHisThrGluSer 1291
```

```
Qy        733 ATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db       1292 GlyGluPheSerArgGlnLeuAspGluLysAspAlaMetVal---------SerGlnLeu 1308

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG------TCTATAATAGAA 843
Db       1309 SerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLysArgGlnLeuGluGlu 1328

Qy        844 AAAACAAAT---------------------------------------------------- 852
Db       1329 GluThrLysAlaLysSerThrLeuAlaHisAlaLeuGlnSerAlaArgHisAspCysAsp 1348

Qy        853 ------AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA--- 903
Db       1349 LeuLeuArgGluGlnTyrGluGluGluGlnGluAlaLysAlaGluLeuGlnArgGlyMet 1368

Qy        904 ----------------------------CATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db       1369 SerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGln 1388

Qy        934 AAACCAGAACAAAAAGAAGAT--------------------AAATCTGCTGACATACAA 972
Db       1389 ArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnAspAlaGlu 1408

Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
Db       1409 GluHis---ValGluAlaValAsnSerLysCysAlaSer---------LeuGluLysThr 1424

Qy       1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------ 1086
Db       1425 LysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGluArgSerAsnAla 1444

Qy       1087 ----------------------------------GATGAAGACTTAGACGAATTTAAG 1110
Db       1445 AlaCysIleAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuAlaGluTrpLys 1464

Qy       1111 CCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------- 1152
Db       1465 Gln------LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArg 1482

Qy       1153 ----------------------------GGAATTTATAAAGAA------------CTA 1170
Db       1483 SerLeuSerThrGluLeuPheLysValLysAsnAlaTyrGluGluSerLeuAspHisLeu 1502

Qy       1171 GAAGATTTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGATGAAGGA 1218
Db       1503 GluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGln 1522

Qy       1219 ATAGAAAAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATAT 1275
Db       1523 IleAlaGluGlyGlyLysHisIleHisGluLeuGluLysValLysLysGlnLeuAspHis 1542

Qy       1276 GAAAAAACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332
Db       1543 GluLysSerGluLeuGlnThrSerLeuGluGluAlaGluAlaSerLeuGluHisGluGlu 1562

Qy       1333 ATTAAA----------------------------------AAATATAAAAATGATAAGCAG 1359
Db       1563 GlyLysIleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluIleAspArgLys 1582

Qy       1360 GTTAATAAGGAAAAGGAA----------------------------AAATTCATAAAA 1389
Db       1583 IleAlaGluLysAspGluGluLeuAspGlnLeuLysArgAsnHisLeuArgValValGlu 1602

Qy       1390 TCATTGTTTCATATATTTGACGGAGAC------------AATGAAATTTTACAGATCGTG 1437
Db       1603 SerMetGlnSerThrLeuAspAlaGluIleArgSerArgAsnAspAlaLeuArgIleLys 1622

Qy       1438 GATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
Db       1623 LysLysMetGluGlyAspLeuAsnGluMetGluIleGlnLeu 1636

RESULT 30
```

```
US-09-538-092-915
; Sequence 915, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 915
;   LENGTH: 1939
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number P12882
US-09-538-092-915

Alignment Scores:
Pred. No.:                1.01e-16      Length:          1939
Score:                    303.00        Matches:         159
Percent Similarity:       41.68%        Conservative:    124
Best Local Similarity:    23.42%        Mismatches:      162
Query Match:              12.18%        Indels:          234
DB:                       2             Gaps:            40

US-09-837-344-45 (1-1482) x US-09-538-092-915 (1-1939)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        931 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 950

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
Db        951 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 970

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        971 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 990

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        991 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1010

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1011 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysIleLys 1030

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db       1031 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysIleArgMet 1050

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1051 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerAla 1070

Qy        268 AGCGATTTAGAACAAGATAGA---CTTGCTAAAGAAAAGTTACAAGAGCAG--------- 315
Db       1071 MetAspIleGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysGluPheGlu 1090

Qy        316 ------------CAAAGCGATTTAGAACAAGAG----------------------AGA 339
Db       1091 MetSerGlyLeuGlnSerLysIleGluAspGluGlnAlaLeuGlyMetGlnLeuGlnLys 1110

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
```



Search completed: December 21, 2005, 09:38:26
Job time : 125.334 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:23 ; Search time 234.352 Seconds
(without alignments)
5284.546 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22847/app_query.fasta_1.3740
-DB=Published_Applications_AA_Main -QFMT=fastan -SUFFIX=rapbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_781_@runat_21122005_093214_22847
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2403.5	96.6	1909	5	US-10-732-923-3341	Sequence 3341, Ap
2	1546	62.2	316	3	US-09-837-344-31	Sequence 31, Appl
3	1546	62.2	316	3	US-09-900-963-31	Sequence 31, Appl
4	600	24.1	117	3	US-09-837-344-20	Sequence 20, Appl
5	600	24.1	117	3	US-09-900-963-20	Sequence 20, Appl
6	551	22.2	107	3	US-09-837-344-19	Sequence 19, Appl
7	551	22.2	107	3	US-09-900-963-19	Sequence 19, Appl
8	382.5	15.4	1407	5	US-10-732-923-3359	Sequence 3359, Ap
9	382.5	15.4	1549	5	US-10-732-923-3358	Sequence 3358, Ap
10	381.5	15.3	1003	5	US-10-732-923-3319	Sequence 3319, Ap
c 11	376.5	14.3	419	5	US-10-450-763-47660	Sequence 47660, A
12	374	15.0	1439	5	US-10-754-079-157	Sequence 157, App
13	361	14.5	350	4	US-10-101-487-58	Sequence 58, Appl
14	361	14.5	350	5	US-10-939-988-58	Sequence 58, Appl
15	359	14.4	1162	3	US-09-894-273-2	Sequence 2, Appli
16	359	14.4	1162	4	US-10-294-804-2	Sequence 2, Appli
17	359	14.4	1162	5	US-10-194-046-2	Sequence 2, Appli
18	337	13.6	1527	5	US-10-732-923-3354	Sequence 3354, Ap
19	330	13.3	2274	6	US-11-097-143-2763	Sequence 2763, Ap
20	325	13.1	611	3	US-09-216-393-81	Sequence 81, Appl
21	325	13.1	611	4	US-10-321-856-81	Sequence 81, Appl
22	321	12.9	658	6	US-11-097-143-23688	Sequence 23688, A
23	321	12.9	800	6	US-11-097-143-41169	Sequence 41169, A
24	318	12.8	554	6	US-11-097-143-5154	Sequence 5154, Ap
25	317.5	12.8	1937	5	US-10-723-860-58	Sequence 58, Appl
26	317.5	12.8	1937	5	US-10-723-860-106	Sequence 106, App
27	317.5	12.8	1961	4	US-10-028-248A-103	Sequence 103, App
28	317.5	12.8	1961	4	US-10-107-782-103	Sequence 103, App
29	315.5	12.7	1170	4	US-10-341-434-95	Sequence 95, Appl
30	315.5	12.7	1564	4	US-10-144-198-2	Sequence 2, Appli
31	315.5	12.7	1564	4	US-10-144-198-4	Sequence 4, Appli
32	315	12.7	647	5	US-10-691-672A-3	Sequence 3, Appli
33	315	12.7	1960	4	US-10-236-031B-62	Sequence 62, Appl
34	315	12.7	1960	4	US-10-028-248A-104	Sequence 104, App
35	315	12.7	1960	4	US-10-107-782-104	Sequence 104, App
36	315	12.7	1960	5	US-10-696-909A-49	Sequence 49, Appl
37	315	12.7	1960	6	US-11-009-554-27	Sequence 27, Appl
38	315	12.7	2143	5	US-10-450-763-32075	Sequence 32075, A
39	313.5	12.6	2056	6	US-11-097-143-4824	Sequence 4824, Ap
40	312.5	12.6	1959	4	US-10-028-248A-36	Sequence 36, Appl
41	312.5	12.6	1959	4	US-10-107-782-36	Sequence 36, Appl
42	311	12.5	240	4	US-10-101-487-75	Sequence 75, Appl
43	311	12.5	240	5	US-10-939-988-75	Sequence 75, Appl
44	310.5	12.5	1639	6	US-11-097-143-4635	Sequence 4635, Ap
45	310.5	12.5	1959	4	US-10-028-248A-106	Sequence 106, App

ALIGNMENTS

RESULT 1
US-10-732-923-3341
; Sequence 3341, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3341
; LENGTH: 1909
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-3341

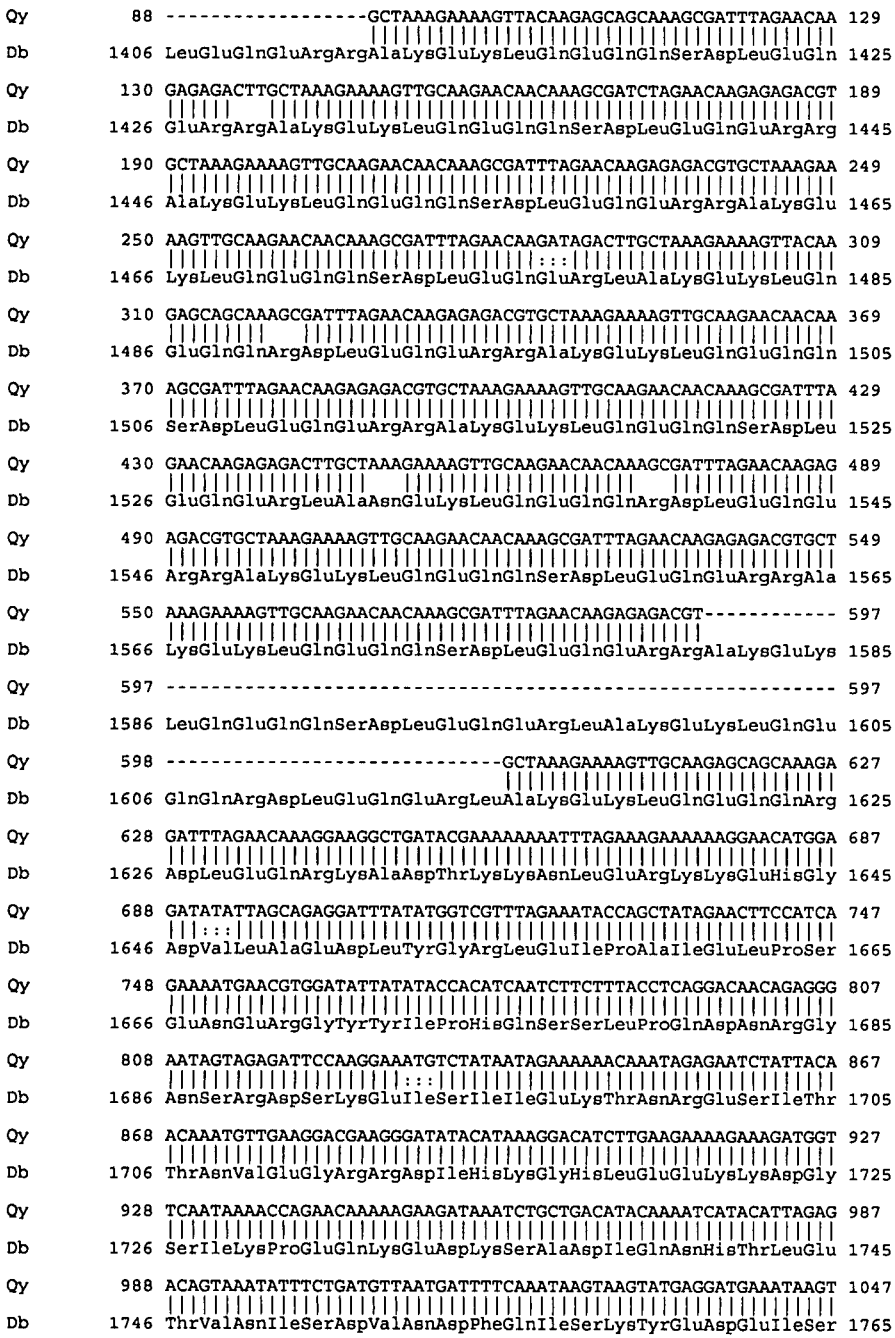
Alignment Scores:

Pred. No.:	8.98e-154	Length:	1909
Score:	2403.50	Matches:	486
Percent Similarity:	89.89%	Conservative:	3
Best Local Similarity:	89.34%	Mismatches:	4
Query Match:	96.64%	Indels:	51
DB:	5	Gaps:	2

US-09-837-344-45 (1-1482) x US-10-732-923-3341 (1-1909)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT---------------------------------- 87
              |||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405
```



RESULT 2
US-09-837-344-31
; Sequence 31, Application US/09837344
; Patent No. US20020041882A1
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/837,344
; FILING DATE: 19-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/462,625
; FILING DATE: <Unknown>
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:

```
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-837-344-31

Alignment Scores:
Pred. No.:              4.77e-96        Length:         316
Score:                  1546.00         Matches:        313
Percent Similarity:     99.37%          Conservative:   1
Best Local Similarity:  99.05%          Mismatches:     2
Query Match:            62.16%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-45 (1-1482) x US-09-837-344-31 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 3
US-09-900-963-31
; Sequence 31, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;         APPLICANT: GUERIN-MARCHAND, Claudine
;                    DRUILHE, Pierre
;         TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                             HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                             OF STIMULATING THE T LYMPHOCYTES
;         NUMBER OF SEQUENCES: 46
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 31:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 316 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;         MOLECULE TYPE: peptide
;         PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-900-963-31

Alignment Scores:
Pred. No.:              4.77e-96        Length:         316
Score:                  1546.00         Matches:        313
Percent Similarity:     99.37%          Conservative:   1
Best Local Similarity:  99.05%          Mismatches:     2
Query Match:            62.16%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-45 (1-1482) x US-09-900-963-31 (1-316)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
```

```
Db          1 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 20

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 40

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 60

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 80

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGlu 100

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSer 120

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGlu 140

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArg 160

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLys 180

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 200

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArg 220

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAla 240

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuPro 260

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        261 GlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsn 280

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 ArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGlu 300

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 316

RESULT 4
US-09-837-344-20
; Sequence 20, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 20:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 117 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-837-344-20

Alignment Scores:
Pred. No.:              2.24e-32        Length:         117
Score:                  600.00          Matches:        116
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.15%          Mismatches:     0
Query Match:            24.13%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-45 (1-1482) x US-09-837-344-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 5
US-09-900-963-20
; Sequence 20, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 20:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 117 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-900-963-20

Alignment Scores:
Pred. No.:              2.24e-32      Length:         117
Score:                  600.00        Matches:        116
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  99.15%        Mismatches:     0
Query Match:            24.13%        Indels:         0
DB:                     3             Gaps:           0

US-09-837-344-45 (1-1482) x US-09-900-963-20 (1-117)

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 20

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 ArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 40

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         41 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 60

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         61 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluLysThr 80

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 100

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 117

RESULT 6
US-09-837-344-19
; Sequence 19, Application US/09837344
; Patent No. US20020041882A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/837,344
;            FILING DATE: 19-Apr-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/462,625
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: FR 91 01286
;            FILING DATE: 05-FEB-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-078
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 19:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 107 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-837-344-19

Alignment Scores:
Pred. No.:              4.45e-29      Length:         107
Score:                  551.00        Matches:        106
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  99.07%        Mismatches:     0
Query Match:            22.16%        Indels:         0
DB:                     3             Gaps:           0
```

```
US-09-837-344-45 (1-1482) x US-09-837-344-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 7
US-09-900-963-19
; Sequence 19, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 19:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 107 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-900-963-19

Alignment Scores:
Pred. No.:              4.45e-29        Length:         107
Score:                  551.00          Matches:        106
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            22.16%          Indels:         0
DB:                     3               Gaps:           0

US-09-837-344-45 (1-1482) x US-09-900-963-19 (1-107)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSer 107

RESULT 8
US-10-732-923-3359
; Sequence 3359, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3359
;   LENGTH: 1407
;   TYPE: PRT
;   ORGANISM: Oryctolagus cuniculus
US-10-732-923-3359

Alignment Scores:
Pred. No.:              8.99e-18        Length:         1407
Score:                  382.50          Matches:        127
Percent Similarity:     45.84%          Conservative:   132
Best Local Similarity:  22.48%          Mismatches:     167
Query Match:            15.38%          Indels:         139
DB:                     5               Gaps:           20

US-09-837-344-45 (1-1482) x US-10-732-923-3359 (1-1407)

Qy          1 CAAGAACAACAAAGCGATCTA---------GAACAAGAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||:::   :::|||         |||:::::|||:::   :::|||:::
Db        140 GlnGluGlnGluArgGluLeuAlaGluGluGluGluGlnArgLysLysArgGluArgPhe 159

Qy         52 CAAGAA------------CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG 99
              ::::::            |||      |||   ||||||   ||||||   :::::::::
```

```
Db        160 GluGlnHisTyrSerArgGlnTyrArgAspLysGluGln---ArgLeuGlnArgGlnGlu 178

Qy        100 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 159
              |||:::|||:::::::::      |||:::|||:::|||   :::   |||   ::::::
Db        179 LeuGluGluArgArgAla------GluGluGluGlnLeuArgArgArgLysGlyArgAsp 196

Qy        160 CAACAAAGCGATCTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAA 210
                 :::     :::|||:::|||         ||||||   :::::::::|||:::
Db        197 AlaGluGluPheIleGluGluGluGlnLeuArgArgArgGluGlnGlnGluLeuLysArg 216

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG--------------TTG 255
              :::      :::   ||||||:::||||||   :::   :::                |||
Db        217 GluLeuArgGluGluGluGlnGlnArgArgGluArgArgGluGlnHisGluArgAlaLeu 236

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              ||||||::::::      |||   :::   |||      :::|||:::   :::||||||
Db        237 GlnGluGluGluGluGlnLeuLeuArgGlnArgArgTrpArgGluGluProArgGluGln 256

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
              |||            |||   ||||||      |||:::::::::|||::::::
Db        257 Gln-----------GlnLeuArgArgGluLeuGluGluIleArgGluArgGluGlnArg 272

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
              ||||||||||||   |||  :::::::::|||:::   :::|||      |||||||||
Db        273 LeuGluGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnArg---LeuGluGln 291

Qy        436 GAG--------------------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              |||                    |||         |||:::::::::|||::::::
Db        292 GluGluArgArgGluGlnGlnLeuArgArgGluLeuGluGluIleArgGluArgGluGln 311

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
              |||||||||||   |||  :::::::::|||::::::::::::      :::
Db        312 ArgLeuGluGlnGluGluArgArgGluGlnArgLeuGluGlnGluGluArgArgGluGln 331

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 594
              |||   :::|||   |||:::::::::|||::::::     ||||||||||||
Db        332 GlnLeuLysArgGluLeuGluGluIleArgGluArgGluGlnArgLeuGluGlnGluGlu 351

Qy        595 CGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||   ::::::   |||   |||:::   |||
Db        352 ArgArgGluGlnLeuLeuAlaGluGluValArg-------------------------- 362

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                    |||:::   :::|||   |||:::   |||
Db        363 -----------GluGlnAlaArgGluArgGlyGluSerLeuThrArgArgTrpGlnArg 378

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
              :::||||||           |||      :::         :::   |||      |||
Db        379 GlnLeuGlu---------SerGluAlaGlyAlaArgGlnSerLysValTyrSerArgPro 395

Qy        775 ---------CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825
                     :::   ||||||   ||||||                |||      :::
Db        396 ArgArgGlnGluGluGlnSerLeuArgGlnAspGlnGluArgArgGlnArgGlnGluArg 415

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGA---------------------------- 855
              |||      :::   |||:::      |||
Db        416 GluArgGluLeuGluGluGlnAlaArgArgGlnGlnGlnTrpGlnAlaGluGluGluSer 435

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT------ 909
              |||                  :::      |||      :::   :::      |||
Db        436 GluArgArgArgGlnArgLeuSerAlaArgProSerLeuArgGluArgGlnLeuArgAla 455

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAA------------CCAGAACAAAAAGAAGATAAA 957
              ||||||:::::::::         :::               ||||||:::   ::::::
Db        456 GluGluArgGlnGluGlnGluGlnArgPheArgGluGluGluGluGlnArgArgGluArg 475

Qy        958 TCTGCTGACATACAA---------------------------------------------- 972
                  ::::::|||
Db        476 ArgGlnGluLeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGln 495

Qy        972 ---------------------------------------------------------- 972

Db        496 GlnLeuGlnGluGluAspSerPheGlnGluAspArgGluArgArgArgArgGlnGlnGlu 515
```

```
Qy        973 ----------------------------------------------------AATCAT 978
                                                                     |||
Db        516 GlnArgProGlyGlnThrTrpArgTrpGlnLeuGlnGluGluAlaGlnArgArgArgHis 535

Qy        979 ACATTA--------------GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
              ||||||              :::         :::   :::      :::   |||
Db        536 ThrLeuTyrAlaLysProGlyGlnGlnGluGlnLeuArgGluGluGluGluLeuGlnArg 555

Qy        1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATAT---GACGATTCATTAATAGATGAAGAA 1080
               |||          |||          ||||||  ::::::      :::      |||
Db        556 GluLysArgArgGlnGluArgGluArgGluTyrArgGluGluGluLysLeuGlnArgGlu 575

Qy        1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
               ||||||::::::                                          :::
Db        576 GluAspGluLysArg-----------------------------------------ArgArg 582

Qy        1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
               :::|||             |||:::|||||||||:::|||   ::::::   :::
Db        583 GlnGluArgGluArgGlnTyrArgGluLeuGluGluLeuArgGlnGluGluGlnLeuArg 602

Qy        1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
               |||                ||:::      :::|||      ||||||:::::::::
Db        603 AspArgLysLeuArgGluGluGluGlnLeuLeuGlnGluArgGluGluGluArgLeuArg 622

Qy        1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT------TTTAAACCAAATGATAAA 1314
               :::   ::::::      |||   ::::::              :::      ::::::
Db        623 ArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgGlnGluGluGln 642

Qy        1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
               |||      |||      :::|||   :::   ::::::|||:::   :::   :::
Db        643 GluLeuArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgArgGlu 662

Qy        1375 GAAAAATTCATAAAA 1389
               |||:::   ::::::
Db        663 GluGlnGluLeuArg 667
```

RESULT 9
US-10-732-923-3358
; Sequence 3358, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3358
; LENGTH: 1549
; TYPE: PRT
; ORGANISM: Ovis aries
US-10-732-923-3358

Alignment Scores:

Pred. No.:	8.96e-18	Length:	1549
Score:	382.50	Matches:	133
Percent Similarity:	43.59%	Conservative:	129
Best Local Similarity:	22.13%	Mismatches:	194
Query Match:	15.38%	Indels:	145
DB:	5	Gaps:	18

US-09-837-344-45 (1-1482) x US-10-732-923-3358 (1-1549)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::   |||:::     :::||||||:::::::|||   :::   :::|||:::|||:::
Db        197 GluArgGlnArgLeuGlnValGluGlnGlnGlnArgGlnArgGlyGluLeuArgGluArg 216

Qy         61 CAAAGCGATTTA---------------GAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
              |||   ::::::               |||          |||   :::|||:::|||:::
Db        217 GlnGluGluValGlnLeuGlnLysArgGluThrGlnGluLeuGlnArgGluArgLeuGlu 236
```

```
Qy      106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
            |||:::|||         :::|||:::|||      :::|||:::|||   ||||||:::
Db      237 GluGluGlnGlnLeuGlnLysGlnLysArgGlyLeuGluGluArgLeuLeuGluGlnGlu 256

Qy      166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
                  ::::::   ||||||   :::::::::   :::|||||||||
Db      257 ArgArgGluGlnGluLeuArgArgLysGluGlnGluArgArgGluGlnGlnLeuArgGln 276

Qy      226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------------------- 264
            |||||||||         :::|||::::::   |||:::
Db      277 GluGlnGluGluAlaThrGlnGluGluIleSerGluArgGlyGluSerArgThrSerArg 296

Qy      265 ---CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG------------------ 303
               |||      ||||||   :::      |||:::::::::
Db      297 CysGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnArgLysValTyrSerArgPro 316

Qy      304 ---TTACAAGAGCAGCAAAGC-----------------------GATTTAGAACAAGAG 336
                |||||||||||||||                       :::   :::||||||
Db      317 HisArgGlnGluGlnGlnSerArgArgGlnGluGlnGluLeuGluArgGlnGlnGlu 336

Qy      337 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------------TTAGAA 381
            ::::::      |||::::::|||      |||   |||                |||:::
Db      337 GlnGlnIleSerGluGluValGlnSerLeuGlnGluAspGlnGlyArgGlnArgLeuLys 356

Qy      382 CAAGAGAGACGT---------------------------------------------- 393
            ||||||:::|||
Db      357 GlnGluGlnArgTyrAspGlnAsnTrpArgTrpGlnLeuGluGluGluSerGlnArgArg 376

Qy      394 ---------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
                          ||||||      :::||||||      :::   ||||||
Db      377 ArgTyrThrLeuTyrAlaLysProAlaGlnArgGluGlnValArgGluGluGluGlnLeu 396

Qy      439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA------ 492
            ||||||   :::||||||||||||   ::::::         |||:::||||||
Db      397 ArgLeuLysGluGluLysLeuGlnArgGluLysArgArgGlnGluArgGluArgGlnTyr 416

Qy      493 -----------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                             |||   :::|||:::||||||   ::::::      |||:::
Db      417 ArgGluValGluLeuGlnArgGluGluGluArgLeuGlnArgGluGluGluGlnLeuGln 436

Qy      535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA-------------------------- 567
            :::|||   |||   |||   :::   ||||||
Db      437 ArgGluGluArgGluLysArgArgArgGlnGluArgGluLysGlnTyrLeuGluLysVal 456

Qy      568 ---------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
                     ::::::      |||:::::|||   |||   |||   :::   :::
Db      457 GluLeuTrpGluGluGluGlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArg 476

Qy      619 CAGCAAAGAGATTTAGAA-----------------------------CAAAGGAAGGCT 648
            :::::::::   ||||||                             ||||||:::
Db      477 GluLysGlnTyrLeuGluLysValGluLeuArgGluGluGluGlnLeuGlnArgGlnGlu 496

Qy      649 GATACGAAAAAAAATTTAGAAAGAAAAAAG---------GAACATGGAGATATATTAGCA 699
                ||:::      ||||||:::::      |||      ::::::
Db      497 ArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGlu 516

Qy      700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
            |||:::          |||   |||           :::          |||   ||||||
Db      517 GluGluGlnLeuGlnArgGluGlu------ArgGluLysArgArgGlnGluArgGluArg 534

Qy      760 GGATATTAT------------ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
               |||               :::      :::      |||   :::      |||
Db      535 GlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGlu 554

Qy      808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA--------- 858
                |||      :::|||         :::||||||         :::
Db      555 LysArgArgGlnGluArgGluLysGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 574

Qy      859 ------TCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                                  :::   |||::::::      |||   :::||||||
Db      575 GlnLeuGlnArgGlnGluArgGlnLysArgArgGlnGluArgGluLysGlnTyrLeuGlu 594

Qy      913 GAA---------AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
            :::         :::::::::      ::::::   :::::::::|||   :::
Db      595 LysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGluLysArgArgGln 614

Qy      964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA--- 1020
            :::   :::         ||||||   |||   :::   :::            |||
Db      615 GluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnValGlnArg 634

Qy      1021 ------------------------------------------------------ATAAGT 1026
                                                                     :::
Db      635 GlnGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysGluLeuGln 654

Qy      1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
             :::   |||:::   :::      |||      :::   |||:::   ||||||   :::
Db      655 ArgGlnGluGluArgLeuGlnGluGlu------GluGlnLeuLeuArgGluGluArgGlu 672

Qy      1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                     :::      :::      ::::::                |||:::||||||
Db      673 LysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 692

Qy      1147 AACATA---------------------------------------GGAATTTATAAAGAA 1167
                ::::                                          |||   |||
Db      693 GlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGlu 712

Qy      1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
             |||:::|||      :::   |||   |||            |||   :::::::::
Db      713 LysGluGluLeuGlnArgGlnGluGluArgLeuGlnArgGluLysGlnLeuGlnArg 732

Qy      1228 TCATCAGAAGAA----------------------------TTATCTGAAGAAAAAATA 1257
                  |||                                 |||   ||||||
Db      733 GluAspArgGluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGluLeuGln 752

Qy      1258 AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
             ::::::      :::   ::::::   |||          :::      :::|||
Db      753 GlnGluGluAspArgLeuGlnArgGluLysGlnLeuLeuArgGluAspArgGluLysArg 772

Qy      1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
             |||   |||   ::::::      :::   :::::::|||:::   :::|||||||
Db      773 GlnTyrLeuGluLysValGluLeuGlnArgGluGluGluGlnLeuGlnArgGluLysArg 792

Qy      1378 AAA 1380
             :::
Db      793 Arg 793
```

RESULT 10
US-10-732-923-3319
; Sequence 3319, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3319
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-732-923-3319

Alignment Scores:

Pred. No.:	1.06e-17	Length:	1003
Score:	381.50	Matches:	101
Percent Similarity:	53.41%	Conservative:	126
Best Local Similarity:	23.76%	Mismatches:	105
Query Match:	15.34%	Indels:	93
DB:	5	Gaps:	15

US-09-837-344-45 (1-1482) x US-10-732-923-3319 (1-1003)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db        440 GlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProLeuGlnGluProGln 459

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
Db        460 GlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGln 479

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
Db        480 GlnGlnGluProGlnGlnGlnGluProGlnGlnGlnAspGluGlnGlnGlnAspGluGln 499

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 243
Db        500 GlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAsp---GluGlnGluGlnGlnAsp 518

Qy        244 AAAGAAAAGTTGCAAGAACAACAAAGCGAT------TTAGAACAAGATAGACTTGCTAAA 297
Db        519 GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGluGln 538

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db        539 GlnGluGluGlnGluGlnGlnGluGluGlnGluGlnGlnGluGluGlnGluGlnGluLeu 558

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db        559 GluGluGlnGluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGln 575

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
Db        576 GluGlnGluLeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGlu 592

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
Db        593 LeuGluGluGln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu 609

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db        610 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 624

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db        625 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 643

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
Db        644 GluGln---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGlu 662

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db        663 LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGlu---------------------- 676

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
Db        676 ------------------------------------------------------------ 676

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        677 LeuGluGluGlnGluGlnGlu---------------------------------------- 683

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db        684 ---------LeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluGlnGlu--- 699

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
Db        700 ------------------------------------LeuGluGluValGluGluGln 706

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
Db        707 GluGlnGluGlnGluGluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGlu 726

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
Db        727 GluGlnGluGluGlnGluLeuGluGluValGlu----------------------Glu 738

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA--- 1194
Db        739 GlnGluGlu--------------GlnGluLeuGluGluValGluGluGlnGluGluGln 753

Qy       1195 AATTTAGATGATTTAGATGAA-----------GGAATAGAAAAATCATCAGAAGAATTA 1242
Db        754 GluLeuGluGluValGluGluGlnGluGlnGlyValGluGlnGlnGluGlnGluThr 773

Qy       1243 TCTGAAGAAAAAATA 1257
Db        774 ValGluGluProIle 778
```

RESULT 11
US-10-450-763-47660
; Sequence 47660, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 47660
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (362)..(408)
; OTHER INFORMATION: Vertebrate metallothioneins proteins domain identified by
; OTHER INFORMATION: eMATRIX, accession number BL00203, p-value=2.484e-11, raw score o
; OTHER INFORMATION: 13.94
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(419)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-47660

Alignment Scores:

Pred. No.:	2.36e-17	Length:	419
Score:	376.50	Matches:	149
Percent Similarity:	44.34%	Conservative:	35
Best Local Similarity:	35.90%	Mismatches:	76
Query Match:	14.25%	Indels:	155
DB:	5	Gaps:	25

US-09-837-344-45 (1-1482) x US-10-450-763-47660 (1-419)

```
Qy       1217 CCTTCATCTAAATCATCTAAATTTTCATTTTTCTCTATCAAATCTTCTAGTTCTTTATAA 1158
Db        136 ProAlaProAlaSerSerProCysSer---ArgSerLeuSerSerSerArgSerLeu--- 153

Qy       1157 ATTCCTATGTTTTCTTCATCTTGGAAA-----------------TTGTCATATTGCACA 1116
Db        154 -----------SerSerSerCysArgValSerCysCysHisSerLeuSerSerCysCys 169

Qy       1115 ATAGGCTTAAATTCGTCTAAGTCTTCATCATCTTCTTCTTCATCTATTAATGAATCGTCA 1056
Db        170 ArgSerLeuSerCysSerArgSerSerSerThrSerSerArgSerSerSerThrSerSer 189

Qy       1055 TATTCAGCACTTATTTCATCCTCATACTTACTTATTTGAAAATCATTAACATCAGAAATA 996
```

```
Db        190 SerSerLeuSerLeuSerSerSerLeu-------------------------------- 198

Qy        995 TTTACTGTCTCTAATGTATGATTTTGTATGTCAGCAGATTTATCTTCTTTTTGTTCTGGT 936
                                  |||   |||      ||||||      ||||||
Db        199 ---------------------SerCysSerSerCysArgLeuSer------CysSer--- 208

Qy        935 TTTATTGAACCATCTTTCTTTTCTTCAAGATGTCCTTTATGTATATCCCTTCGTCCTTCA 876

Db        208 ------------------------------------------------------------ 208

Qy        875 ACATTTGTTGTAATAGATTCTCTATTTGTTTTTTCTATTATAGACATTTCCTTGGAATCT 816

Db        208 ------------------------------------------------------------ 208

Qy        815 CTACTATTCCCTCTGTTGTCCTGAGGTAAAGAAGATTGATGTGGTATATAATATCCACGT 756

Db        209 ------------------------------------------------------------Asn 209

Qy        755 TCATTTTCTGATGGAAGTTCTATAGCTGGTATTTCTAAACGACCATATAAATCCTCTGCT 696
              |||||||||      ||||||         :::|||
Db        210 SerPheSerThrSerSerSerSerLeuSerLeuSer----------------------- 221

Qy        695 AATATATCTCCATGTTCCTTTTTTCTTTCTAAATTTTTTTTCGTATCAGCCTTCCTTTGT 636
                                                  |||:::      |||
Db        222 ---------------------------------------------SerSerLeuSerCys 226

Qy        635 TCTAAATCTCTTTGCTGCTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCG 576
              |||   |||   |||   ||||||:::   |||   :::      |||   |||   |||
Db        227 Ser---SerCysCysLeuSerCysSerAsnSerPheSerThrSerSerSerSerLeuSer 245

Qy        575 CTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGT 516
              |||       |||   :::   |||       ||||||||||||||   |||
Db        246 LeuSerSerSerLeuSerCysSerSerCysArgLeuSerCysSerAsnSerPhe------ 263

Qy        515 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAAC 456
                  :::   |||   :::   |||||||||      |||      ||||||   :::
Db        264 ------SerThrSerSerSerSerLeuSerCysProGlySerPheSerCysSerHisSer 281

Qy        455 TTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTA 396
                |||             |||||||||:::||||||      ||||||:::   ||||||
Db        282 LeuSer------------SerCysSerArgSerLeu---SerSerCysSerHisSerLeu 296

Qy        395 GCACGTCTCTCTTGTTCTAAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTC 336
              :::      |||||||||:::||||||      ||||||:::   ||||||:::
Db        297 Ser------SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------ 311

Qy        335 TCTTGTTCTAAATCGCTTTGCTGCTCTTGTAACTTTTCTTTAGCAAGTCTATCTTGTTCT 276
              |||||||||:::||||||      ||||||:::   ||||||:::      ||||||||
Db        312 SerCysSerArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSer 328

Qy        275 AAATCGCTTTGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAAATCGCTT 216
              :::||||||      ||||||:::   ||||||:::      |||||||||:::||||||
Db        329 ArgSerLeu---SerSerCysSerHisSerLeuSer------SerCysSerArgSerLeu 345

Qy        215 ---TGTTGTTCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTTTGTTGT 159
                 |||||||||:::   |||            ||||||||||||||||
Db        346 SerSerCysSerCysSerLeuSer-----------SerCysSerArgSerLeu---Ser 360

Qy        158 TCTTGCAACTTTTCTTTAGCAAGTCTCTCTTGTTCTAAATCGCTTTGCTGCTCTTGTAAC 99
              ||||||:::   ||||||:::      |||||||||   ||||||      ||||||:::
Db        361 SerCysSerHisSerLeuSer------SerCysSerHisSerLeu---SerSerCysSer 377

Qy         98 TTTTCTTTAGCAAGTCTATCTTGTTCTAAATCGCTTTGT-----------------TGT 57
                ||||||:::|||   |||||||||   |||   |||                   |||
Db        378 ArgSerLeuSerSerCysSerCysSerLeuSerSerCysSerArgCysLeuSerSerCys 397

Qy         56 TCTTGCAACTTTTCTTTAGCACGTCTCTCTTGTTCTAGATCGCTT 12
              ||||||:::   |||            ||||||||||||||||
Db        398 SerCysSerLeuSer-----------SerCysSerArgSerLeu 408

RESULT 12
US-10-754-079-157
; Sequence 157, Application US/10754079
; Publication No. US20050008602A1
; GENERAL INFORMATION:
;  APPLICANT: Hirai, Yohei
;  APPLICANT:  Takebe, Kyoko
;  TITLE OF INVENTION: OLIGOPEPTIDES FOR PROMOTING HAIR GROWTH
;  FILE REFERENCE: 467392000222
;  CURRENT APPLICATION NUMBER: US/10/754,079
;  CURRENT FILING DATE:  2004-01-07
;  PRIOR APPLICATION NUMBER: US 10/155,922
;  PRIOR FILING DATE: 2002-05-23
;  PRIOR APPLICATION NUMBER: PCT/JP01/04691
;  PRIOR FILING DATE: 2001-06-04
;  PRIOR APPLICATION NUMBER: JP 2000-166903
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: JP 2001-347340
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: JP 2001-347338
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: JP 2001-371175
;  PRIOR FILING DATE: 2001-12-05
;  PRIOR APPLICATION NUMBER: JP 2001-371336
;  PRIOR FILING DATE: 2001-12-05
;  PRIOR APPLICATION NUMBER: JP 2003-1891
;  PRIOR FILING DATE: 2003-01-08
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157
;   LENGTH: 1439
;   TYPE: PRT
;   ORGANISM: Mus Musculus
US-10-754-079-157

Alignment Scores:
Pred. No.:              3.35e-17        Length:         1439
Score:                  374.00          Matches:        143
Percent Similarity:     48.15%          Conservative:   130
Best Local Similarity:  25.22%          Mismatches:     170
Query Match:            15.04%          Indels:         124
DB:                     5               Gaps:           25

US-09-837-344-45 (1-1482) x US-10-754-079-157 (1-1439)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------------ 48
              :::|||::::::   :::|||||||||||||||||   :::::::::
Db        597 GluGluGluLysArgGluLeuGluGlnGluArgArgArgArgGlnGlnArgAspArgGln 616

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA---------GAA 96
                 |||:::|||:::|||          |||       |||   |||:::          |||
Db        617 IleLeuGluGluGluGlnPheGlnArgGluHisGlnArgGluAlaArgArgArgAspGlu 636

Qy         97 AAGTTACAAGAGCAGCAA---------------------------------------- 114
                    ||||||::::::
Db        637 ThrPheGlnGluGluGluGlnLeuGlnGlyGluSerArgArgArgGlnGlnGluArgGlu 656

Qy        115 ---AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
                    ||||||:::||||||      :::   :::   :::||||||:::
Db        657 GlyLysPheLeuGluGluGluArgGlnLeuArgThrGluArgGluGluGlnArgArgArg 676

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGAT 222
                 ||||||||||||       :::|||:::          |||||||||::::::   :::
Db        677 GlnGluGlnGluArgGluPheGlnGluGluGluGluHisLeuGlnGluArgGluLysGlu 696

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              |||   ||||||          |||   :::   ||||||::::::          |||:::
Db        697 LeuArgGlnGluCysAspArgLysSerArgGluGlnGluArgArgGlnArgGluGlu 716

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
              ::::::|||   :::::::::   ::::::|||:::          ||||||||||||
Db        717 GluGlnLeuArgArgGlnGlu---ArgAspGlnArgPheArgArgGluGlnGluArgHis 735

Qy        340 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 390
                    |||   :::|||:::|||:::::::::   |||      ||||||||||||
Db        736 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 755
```

```
Qy      391 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 441
Db      756 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 775

Qy      442 ------CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 492
Db      776 GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 795

Qy      493 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA--- 543
Db      796 GlnGluArgGluGluGluGlnLeuArgAspArgProPheArgArgGluGlnGluArgArg 815

Qy      544 ------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
Db      816 LeuGluArgGluGluGluGlnLeuArgAspArgProSerArgArgGluGlnGluArgHis 835

Qy      598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db      836 ---GlnGluArgGluGluGluGlnLeuArgAspArgProSerArgArg---GluGlnGlu 853

Qy      658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
Db      854 ArgArgLeuGluArgGluGluGluGln---------LeuArgAspArgSerPheArgArg 870

Qy      718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db      871 GluGln------------GluLeuArgArgAspArgLysPheHisGluGluGluGluArg 886

Qy      778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT------------------ 819
Db      887 ArgGluGluLeuGluGluGluGlnArgGlyGlnGluArgAspArgLeuArgValGluGlu 906

Qy      819 ------------------------------------------------------------ 819
Db      907 GlnLeuArgGlyGlnArgGluGluGluGlnArgArgArgGlnGluCysAspArgLysLeu 926

Qy      820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA--------------TCTATT 864
Db      927 HisArgGluLeuGluValArgGlnGluLeuGluGluGluArgLeuArgAspArgLysLeu 946

Qy      865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db      947 ArgArgGluGlnGluLeuArgArgAsp---ArgLysPheHisGluGluGluGluArg--- 964

Qy      925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT------ 978
Db      965 ---------ArgHisGluGluPheGluGluLysGlnLeuArgLeuGlnGluProAspArg 981

Qy      979 ----------------------ACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
Db      982 ArgPheArgArgGluGlnGluLeuArgGlnGluCysValGluGluGluArgLeuArgAsp 1001

Qy     1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGAT------------ 1062
Db     1002 SerLysIleArgArg---GluGlnGluLeuArgArgGluArgGluGluGluArgLeuArg 1020

Qy     1063 ----------------------------------TCATTAATAGATGAAGAAGAAGAT 1086
Db     1021 AspArgLysIleArgArgAspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArg 1040

Qy     1087 GATGAAGACTTAGAC---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
Db     1041 ArgGlnGluLeuAspArgLysPheArgGluGluGlnGluLeuAspGlnGluLeuGluGlu 1060

Qy     1144 GAAAACATA---------GGAATTTATAAAGAACTAGAAGATTTGATAGAG--------- 1185
Db     1061 GluArgLeuArgAspArgLysIleArgArgGluGlnGluLeuArgArgGluGlnGluLeu 1080

Qy     1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
Db     1081 ArgArgGluGlnGluPheArgArgGluGlnGluLeuArgArgGluGlnGluPheArgArg 1100

Qy     1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
Db     1101 GluGlnGluLeuArgGlnGluArgGluGluGluArgLeuArgAspArgLysIleArgArg 1120

Qy     1306 AATGAT---------AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
Db     1121 AspGlnGluLeuArgGlnGlyLeuGluGluGluGlnLeuArgArgGlnGluArgAspArg 1140

Qy     1357 CAGGTTAATAAGGAAAAGGAA 1377
Db     1141 LysPheArgGluGluGlnGlu 1147
```

RESULT 13
US-10-101-487-58
; Sequence 58, Application US/10101487
; Publication No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-58

Alignment Scores:

Pred. No.:	2.62e-16	Length:	350
Score:	361.00	Matches:	80
Percent Similarity:	50.72%	Conservative:	132
Best Local Similarity:	19.14%	Mismatches:	138
Query Match:	14.52%	Indels:	68
DB:	4	Gaps:	4

US-09-837-344-45 (1-1482) x US-10-101-487-58 (1-350)

```
Qy        1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        1 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20

Qy       61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db       21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40

Qy      121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db       41 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 60

Qy      181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db       61 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80

Qy      241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db       81 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100

Qy      301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db      101 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
```

```
Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db        121 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 140

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        141 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        161 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        181 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
Db        201 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db        221 GluGluGluGluGluGlu---------------------------------------- 227

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db        228 --------------GluGluGluGluGluGluGlu----------------------- 234

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
Db        235 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 250

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db        251 GluGluGluGluGluGlu---------------------------------------- 257

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db        258 ---------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 274

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        275 GluGlu---------------------------------------------------- 276

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        277 ------------GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 292

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db        293 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 312

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db        313 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 332

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
Db        333 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 350

RESULT 14
US-10-939-988-58
; Sequence 58, Application US/10939988
; Publication No. US20050118136A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/939,988
;  CURRENT FILING DATE:  2004-09-14
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 350
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-939-988-58

Alignment Scores:
Pred. No.:                2.62e-16        Length:          350
Score:                    361.00          Matches:         80
Percent Similarity:       50.72%          Conservative:    132
Best Local Similarity:    19.14%          Mismatches:      138
Query Match:              14.52%          Indels:          68
DB:                       5               Gaps:            4

US-09-837-344-45 (1-1482) x US-10-939-988-58 (1-350)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db          1 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db         21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db         41 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 60

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db         61 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 80

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db         81 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 100

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db        101 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db        121 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 140

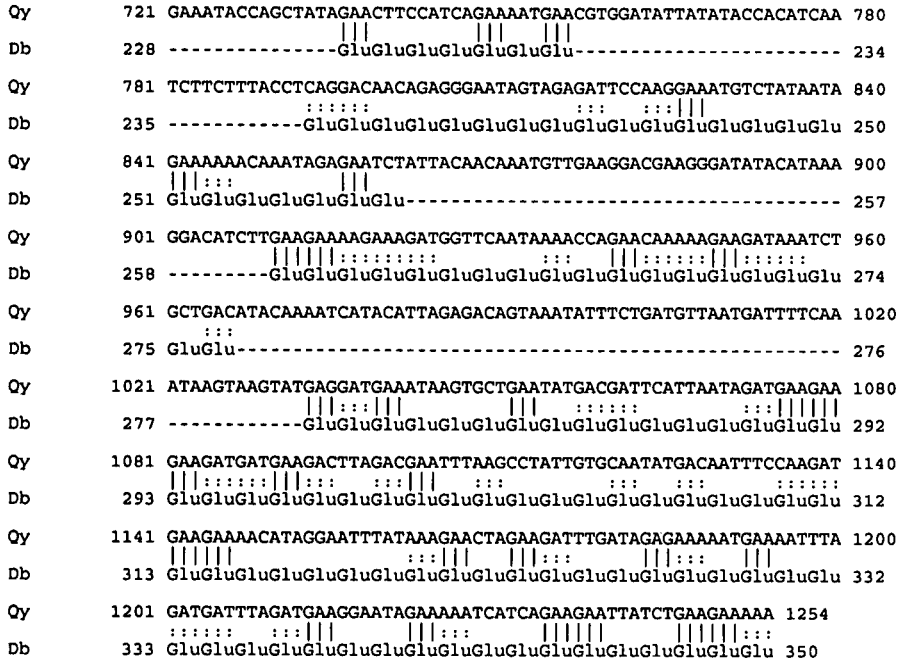
Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        141 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        161 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        181 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
Db        201 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db        221 GluGluGluGluGluGlu---------------------------------------- 227
```

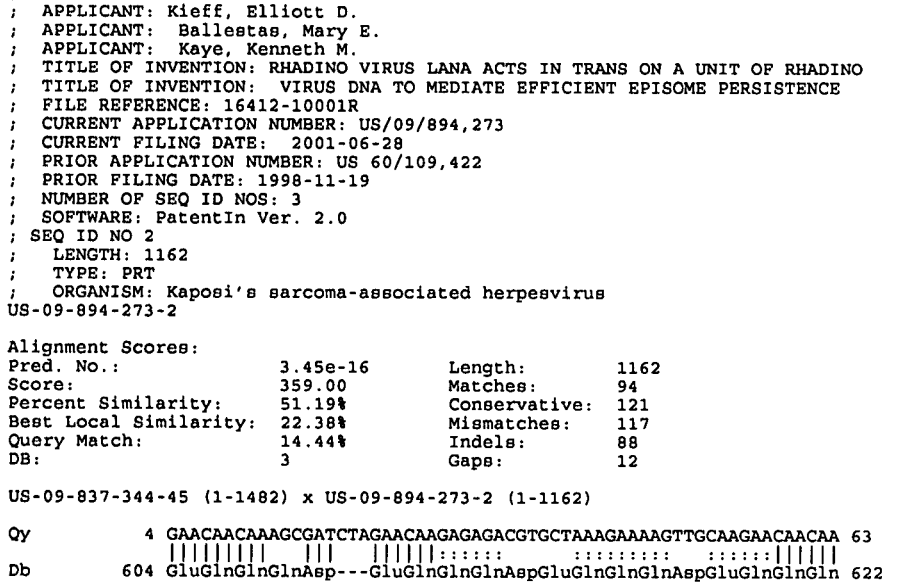


RESULT 15
US-09-894-273-2
; Sequence 2, Application US/09894273
; Publication No. US20040037847A1
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/894,273
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/109,422
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-2

Alignment Scores:

Pred. No.:	3.45e-16	Length:	1162
Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	3	Gaps:	12

US-09-837-344-45 (1-1482) x US-09-894-273-2 (1-1162)



```
Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              ::::::|||               :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                        :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 16
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE: UM-03778
;  CURRENT APPLICATION NUMBER: US/10/294,804
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: US/09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

Alignment Scores:
Pred. No.:              3.45e-16       Length:         1162
Score:                  359.00         Matches:        94
Percent Similarity:     51.19%         Conservative:   121
Best Local Similarity:  22.38%         Mismatches:     117
Query Match:            14.44%         Indels:         88
DB:                     4              Gaps:           12

US-09-837-344-45 (1-1482) x US-10-294-804-2 (1-1162)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||          :::::::::   |||      |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::         ::::::      |||:::|||             :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                |||::::::      |||:::   |||:::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                  :::::::::   ::::::||||||   :::   :::||||||         :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::         ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::         :::::::::   :::::::||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln-----------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                ||||||:::         :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::         :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
                 :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::       :::       :::|||              |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                                 |||   |||
Db        828 --------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                       :::   :::|||   :::
Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              |||:::         |||                  |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::       |||||||||   ::::::         :::   |||:::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017

Db        879 ---------------------------------------------------------- 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::         |||   ::::::            |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                              :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu---------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              ::::::|||               :::|||||||||::::::   |||:::   :::
Db        908 GluGlnGluGln-----------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                        :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 17
US-10-194-046-2
; Sequence 2, Application US/10194046
; Publication No. US20040248081A1
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  TITLE OF INVENTION: LANA Viral Protein Binding Sites
;  FILE REFERENCE: UM-07142
;  CURRENT APPLICATION NUMBER: US/10/194,046
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-194-046-2

Alignment Scores:
Pred. No.:              3.45e-16       Length:         1162
```

Score:	359.00	Matches:	94
Percent Similarity:	51.19%	Conservative:	121
Best Local Similarity:	22.38%	Mismatches:	117
Query Match:	14.44%	Indels:	88
DB:	5	Gaps:	12

US-09-837-344-45 (1-1482) x US-10-194-046-2 (1-1162)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||||||||   |||   ||||||::::::      :::::::::   ::::::||||||
Db        604 GluGlnGlnGlnAsp---GluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 622

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA------GAGCAGCAAAGC 117
                 :::   :::||||||          :::::::::   |||       |||||||||
Db        623 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 642

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              |||   :::|||:::          ::::::     |||:::|||          :::
Db        643 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 662

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 237
                 |||::::::      |||:::   |||:::::::|||         ||||||::::::
Db        663 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGln 682

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
                 :::::::::   ::::::||||||   :::   :::||||||          :::
Db        683 AspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGlnGlnAspGluGlnGluGln 702

Qy        298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
              :::      ||||||||||||   :::   ||||||:::          ::::::
Db        703 GlnAspGluGlnGluGlnGlnAspGluGlnGluGlnGlnAspGluGlnGlnGlnAspGlu 722

Qy        358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
              |||:::||||||   :::   :::|||:::          :::::::::   ::::::|||
Db        723 GlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGlnGlnGlnAspGluGlnGln 742

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||   :::|||:::            :::      ||||||||||||   :::
Db        743 GlnGlnAspGluGlnGlnGln------------GlnAspGluGlnGluGlnGlnGluGlu 758

Qy        478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                 ||||||:::          :::::::::|||:::||||||:::   :::||||||
Db        759 GlnGluGlnGlnGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluAsp 778

Qy        538 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
              :::          :::::::::|||:::||||||:::   :::||||||::::::
Db        779 Gln---------GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGln------ 793

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
              :::::::::|||:::||||||:::::::::||||||:::::::::   :::   :::
Db        794 ---GluGlnGluLeuGluGluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGlu 812

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              :::      :::      :::|||          |||   |||
Db        813 GluGlnGluGlnGluLeuGluGluGlnGluGlnGluLeuGluGlu--------------- 827

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
                                              |||   |||
Db        828 ---------------------------GlnGluGlnGlu--------------------- 831

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
                                                  :::   :::|||      :::
Db        832 ----------------------------------LeuGluGluGlnGluGlnGluLeu 839

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
              |||:::      |||                 |||      ::::::
Db        840 GluGluGlnGluValGluGluGlnGluGlnGluValGluGluGlnGluGlnGluGlnGlu 859

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
              :::      |||||||||   ::::::         :::   |||::::::|||::::::
Db        860 GluGlnGluLeuGluGluValGluGluGlnGluGlnGluGlnGluGluGlnGluGluGln 879

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
```

```
Db        879 ---------------------------------------------------------- 879

Qy       1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAA 1077
              ::::::   :::   |||::::::          |||   ::::::          |||
Db        880 GluLeuGluGluValGluGluGlnGluGluGlnGluLeuGluGluVal---------Glu 896

Qy       1078 GAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA 1137
              |||:::::::::::::::|||:::|||                              :::
Db        897 GluGlnGluGluGlnGluLeuGluGlu--------------------------ValGlu 907

Qy       1138 GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAAT 1197
              :::::::|||                :::|||||||||:::::   |||:::   :::
Db        908 GluGlnGluGln------------GlnGluLeuGluGluValGluGluGlnGluGln--- 922

Qy       1198 TTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATA 1257
                           :::|||:::|||:::      :::|||      ||||||   |||
Db        923 --------------GlnGlyValGluGlnGlnGluGlnGluThrValGluGluProIle 937

RESULT 18
US-10-732-923-3354
; Sequence 3354, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
;  FILE REFERENCE: 38-15(52796)C
;  CURRENT APPLICATION NUMBER: US/10/732,923
;  CURRENT FILING DATE:  2003-12-10
;  PRIOR APPLICATION NUMBER: 10/310,154
;  PRIOR FILING DATE: 2002-12-04
;  NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3354
;   LENGTH: 1527
;   TYPE: PRT
;   ORGANISM: Plasmodium yoelii yoelii
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1527)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3354
```

Pred. No.:	1.03e-14	Length:	1527
Score:	337.00	Matches:	135
Percent Similarity:	45.21%	Conservative:	134
Best Local Similarity:	22.69%	Mismatches:	190
Query Match:	13.55%	Indels:	136
DB:	5	Gaps:	23

US-09-837-344-45 (1-1482) x US-10-732-923-3354 (1-1527)

```
Qy          1 CAAGAACAA------CAAAGCGATCTAGAACAA---GAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||      :::   ::::::::::   |||   :::   |||||||||:::
Db        755 GlnGluGlnMetLeuLysAspGluIleGlnLysIleGluLeuLysThrLysGluLysIle 774

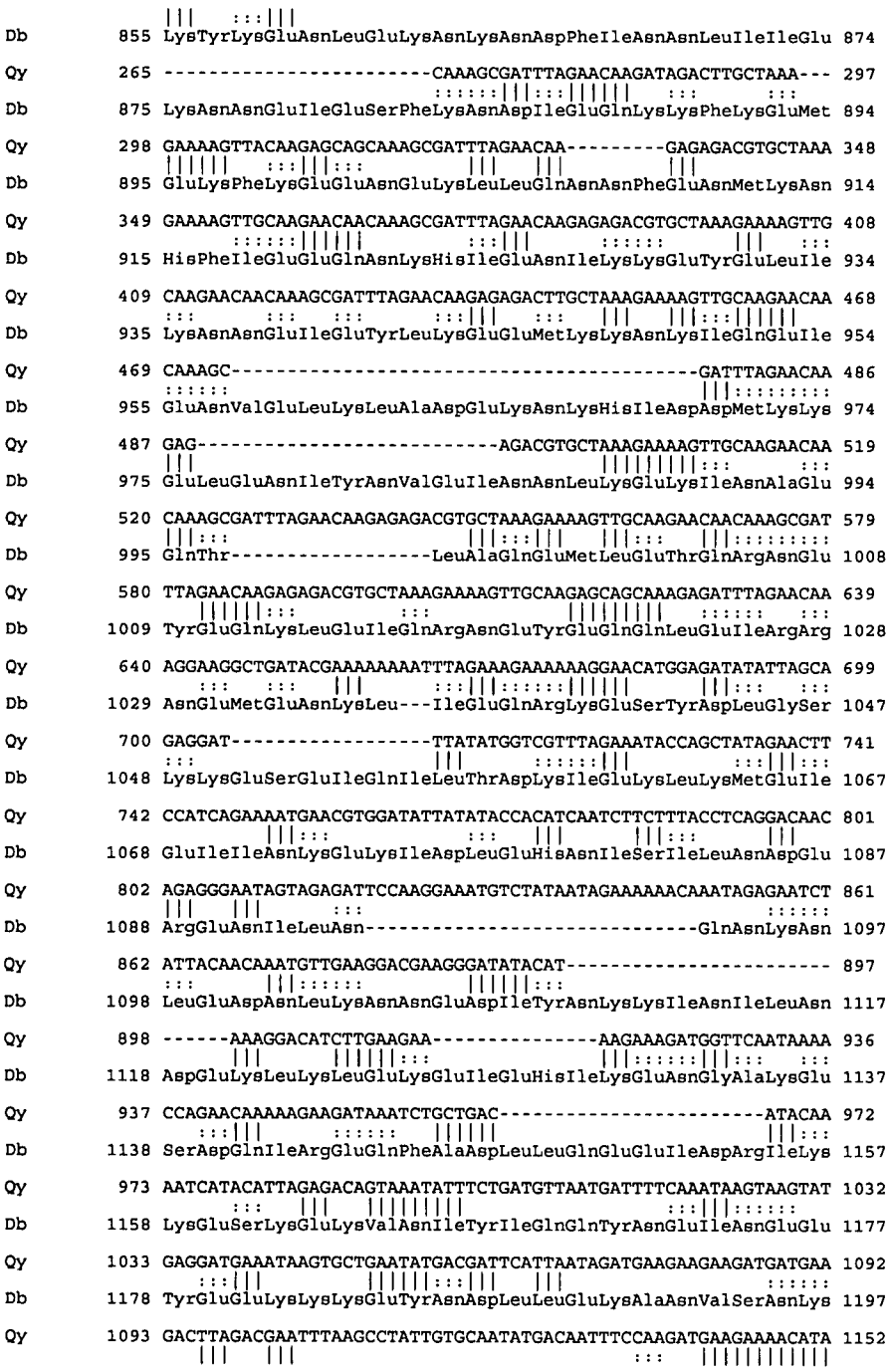
Qy         52 CAAGAACAACAAAGCGATTTAGAACAA-----------------GATAGACTTGCTAAA 93
              :::|||   :::   |||   |||:::                   :::   |||
Db        775 GluGluLeuLysGlnAspPheGluLysThrGlnLysIleAsnMetGluAsnLeuGluMet 794

Qy         94 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG---AGACTTGCTAAAGAAAAG 150
              ||||||                :::::|||||   |||                |||||||||
Db        795 GluLysGluSerPheIleAsnAsnAsnLeuGluAsnGluIleAsnLysMetLysGluLys 814

Qy        151 TTGCAAGAA-----------------CAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              |||:::|||                :::::::::::::::::   :::   :::
Db        815 LeuGluGluLysTyrGluThrGlnIleLysGluThrGluMetLysTyrLysTyrGlnIle 834

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA---GAA 249
              ||||||::::::::::::   :::   :::   ||||||          :::|||   |||
Db        835 LysGluGluIleGluLysThrLysGlnAsnAlaGluGlnAsnPheAsnSerLysPheGlu 854

Qy        250 AAGTTGCAAGAACAA---------------------------------------------- 264
```



RESULT 19
US-11-097-143-2763
; Sequence 2763, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2763
; LENGTH: 2274
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-2763

Alignment Scores:

Pred. No.:	3.02e-14	Length:	2274
Score:	330.00	Matches:	129
Percent Similarity:	43.83%	Conservative:	134
Best Local Similarity:	21.50%	Mismatches:	177
Query Match:	13.27%	Indels:	160
DB:	6	Gaps:	24

US-09-837-344-45 (1-1482) x US-11-097-143-2763 (1-2274)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA---------GAAAAGTTG 51
              :::|||:::::      |||      ||||||      |||         ||||||:::
Db       1385 LysGluArgGluGluSerLeuArgMetGluArgGluGluLysLeuLysGluGluLysIle 1404

Qy         52 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
              :::|||::::::           :::::::::|||   :::||||||::::::|||:::
Db       1405 LysGluLysGlu---------ArgGluGluLysLeuLysGluGluLysIleLysGluLys 1421

Qy        112 CAAAGCGAT------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
              |||   :::      |||:::::::|||:::|||  :::::::::   :::|||:::
Db       1422 GlnArgGluGluLysLeuLysGluGluLysLeuLysGluLysGluArgGluGluArgMet 1441

Qy        166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA--------- 216
                 :::   |||:::|||   :::||||||||||||   :::|||:::::
Db       1442 LysGluLysGluArgGluGluLysAlaLysGluLysGlnArgGluGluLysLeuArgGlu 1461

Qy        217 ---------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                     :::   |||:::|||   :::   ||||||||||||:::|||::::::
Db       1462 GluLysIleLysGluLysGluArgGluGluLysLeuLysGluLysLeuArgGluGluLys 1481

Qy        268 ---AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
                 :::   |||::::::         :::::::::   :::|||:::      :::
Db       1482 IleLysGluLysGluLysGluGluLysLeuArgLysGluArgGluGluLysMetArgGlu 1501

Qy        325 TTAGAACAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAACAA 369
                 |||:::|||   :::   |||                ||||||:::::|||::::::
Db       1502 LysGluArgGluGluLysIleLysGluLysGluArgValGluLysIleLysGluLysGlu 1521

Qy        370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                 :::   :::   ::::::      :::|||||||||:::|||::::::      |||
Db       1522 ArgGluGluLysLeuLysLysGluLysGluGluLysLeuLysGluLysGluGluLeuLeu 1541

Qy        430 GAA--------------------------CAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462
              :::                         :::|||:::|||   :::::::::   |||
Db       1542 LysLysLysGluLysGluGluLysGluArgGluGluLysLeuLysGluLysGluArgGln 1561

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522
              |||:::      :::   |||:::|||   :::   |||   :::   :::|||:::|||
Db       1562 GluLysLeuLysGluLysGluArgGluGluLysLeuLysArgGluThrGluGluArgGln 1581

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGAT 579
                 :::   |||:::|||   :::   |||||||||   :::   |||:::      |||
Db       1582 ArgGluLysGluArgGluGluLysLeuLysGluLysGluArgAlaGluLysLeuLysAsp 1601

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAG---------TTGCAAGAGCAGCAAAGAGAT 630
              ||||||:::|||   :::   |||||||||         |||:::|||:::::::::::::
Db       1602 LeuGluLysGluValLysLeuLysGluLysGluGluGlnLeuLysGluLysGluLysGlu 1621

Qy        631 TTA---------GAACAAAGGAAGGCTGATACGAAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              |||         |||:::::::::   |||      |||      |||:::      |||
Db       1622 LeuLysLeuLysGluLysLysGluLysAspLysValLysGluLysGluLysSerLeuGlu 1641

Qy        682 CATGGAGATATATTA----------------------------------GCAGAGGAT 705
                          :::|||                                 |||.|||
Db       1642 SerGluLysLeuLeuIleSerAlaThrValSerAsnProTrpArgArgValValGluAsp 1661

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATAGAA-------------------------- 738
                    :::|||      ||||||::::::
Db       1662 ThrProProLysLeu------ProAlaValGlnAspTyrProSerLeuGlyLysLysPro 1679

Qy        739 ---------CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                      |||   :::                  :::|||      ::::::
Db       1680 ThrLysAlaSerProGluLysLysArgAspGluLysLeuLeuProGlyLeuThrThrPro 1699

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGAT---------------------------- 819
              |||::::::          |||
Db       1700 ProLysGluValAsnAspThrPheGluAspPheLeuSerGlyLeuLysProLeuGluAla 1719

Qy        820 -----------------TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCT 861
                              :::   :::   :::   |||   :::   :::||||||
Db       1720 LeuProProLeuProAlaLeuGluProLeuGluValLysGluAspSerLysLysGluSer 1739

Qy        862 ATTACA---ACAAATGTTGAAGGA----------------------CGAAGGGATATA 894
```

```
              ::::::      |||   |||                                |||   |||
Db       1740 ValSerLeuIleAsnPheGluSerProLeuGlnGluThrAlaSerIleProArgLysIle 1759

Qy        895 CAT---------AAAGGACATCTTGAAGAAAAG------------AAAGATGGTTCAATA 933
                          :::|||      |||:::                  ||||||:::
Db       1760 SerProProProArgGlyPheThrGluGlnAsnLeuIleLeuAlaLeuCysGlySerLeu 1779

Qy        934 AAACCAGAACAAAAAGAAGATAAA------------TCTGCTGACATACAAAATCAT--- 978
                 |||   ::::::::::::                |||:::::|||
Db       1780 HisTyrGluAsnGluGlnGluArgIleArgGluThrGluAlaGluValGlnProGluVal 1799

Qy        979 --------------------------ACATTAGAGACAGTA----------------- 993
                                       ||||||:::|||:::
Db       1800 ValThrThrProGluProValTyrIleThrLeuAspThrLeuAlaLeuGlnLysSerThr 1819

Qy        994 ------------AATATTTCTGATGTTAATGATTTTCAAATA----------------- 1023
                         |||   |||       :::     :::
Db       1820 SerThrSerAsnAsnSerSerGlySerGluGluIleValValMetGluGluProValLys 1839

Qy       1024 --------------AGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                           |||         ::::::     :::
Db       1840 LysLeuSerLysLysHisLysArgLeuLysGlnLeuGlnLeuGlnGlnGlnGlnGlnGln 1859

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::   |||   |||||||||:::|||                |||         |||
Db       1860 ArgGluArgGluProAspAspGluGluLeuArgProLeuIleSerIleSerMetCysAsp 1879

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                 :::      |||      :::|||   :::
Db       1880 SerGlnLeuAspValProAspIleProVal---------------------------- 1889

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                        :::   |||   |||||||||:::|||:::      |||         |||
Db       1890 ---ArgAlaMetProAspAspAspGluGlyValGluGluGluProGluThrGluProGlu 1908

Qy       1249 GAAAAAATA---------------------------------------------------- 1257
                 :::   :::
Db       1909 GlnSerLeuProGluAspLeuLeuHisPheGlySerSerGlyValAlaThrAsnAlaThr 1928

Qy       1258 ---------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                      :::|||                  |||   |||:::|||
Db       1929 ThrThrAspSerGluGlyProValProAlaThrThrSerAspAspAsnValValTyrGly 1948

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                    |||      :::              |||   |||      |||      :::|||
Db       1949 SerSerSerThrSerAsnAlaAlaProHisLysLeuLysThrLysLysLeuGluHisLys 1968
```

RESULT 20
US-09-216-393-81
; Sequence 81, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393-81

Alignment Scores:

Pred. No.:	6.82e-14	Length:	611
Score:	325.00	Matches:	84
Percent Similarity:	54.84%	Conservative:	103
Best Local Similarity:	24.63%	Mismatches:	100

```
Query Match:            13.07%        Indels:        54
DB:                     3             Gaps:          11

US-09-837-344-45 (1-1482) x US-09-216-393-81 (1-611)

Qy         22 GAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||:::|||||||||             ||||||:::   ::::::::::::   :::
Db        292 GluArgGluArgArgArgAlaGluGluGluLysGluArgGluArgGlnGluGluGluGlu 311

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                 |||:::   |||:::   :::||||||   :::|||:::|||   :::   |||:::
Db        312 ArgGluArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArg 331

Qy        130 GAG------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||      |||:::   :::||||||   ::::::::::::   |||      :::
Db        332 GluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluAspGluArgGluArg 351

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 240
              ||||||   :::|||:::         |||:::   :::   |||:::|||||||||
Db        352 ArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArg 371

Qy        241 ---------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                      |||   :::   :::|||::::::   :::   |||:::   |||:::
Db        372 ValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArgVal 391

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::||||||   ::::::::::::   :::   |||:::|||||||||      |||
Db        392 GluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGlu 411

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-------------- 396
                 :::   :::   :::|||   :::   |||:::||||||
Db        412 GluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGlu 431

Qy        397 ---AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
                  |||   :::   :::|||::::::   :::   |||      |||:::   :::|||
Db        432 GluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGluGlu 451

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
              |||   ::::::::::::   :::      :::   ||||||   :::|||:::   :::
Db        452 LysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGluGluGluLysGlu 471

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAA 567
                 ::::::   :::   |||:::|||||||||        :::||||||   :::|||
Db        472 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGlu 491

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAA 615
                 :::|||   :::   |||:::|||||||||             |||||:::   :::
Db        492 ArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGluArg 511

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
                 :::::::::          |||:::|||:::   :::   :::|||      :::|||:::
Db        512 GlnGluGluGluLysArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGln 531

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
                 :::|||   |||
Db        532 GluGluGluGly------------------------------------------------ 535

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
                       |||   :::||||||             :::             :::
Db        536 ---------ArgGluArgGlnArgGly---------GluGluArgGluGluArgGluArg 549

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAAT 852
                 :::   :::          |||:::   |||          :::   :::::::::   :::
Db        550 GluPheGlnGlnArgGluArgGluLeuLysThrArgLeuValGluLeuGlnArgGluHis 569

Qy        853 AGAGAATCTATTACAACA------AATGTTGAAGGACGAAGGGATATACAT--------- 897
                 ||||||:::   |||           :::|||   |||:::   |||
Db        570 AlaGluSerValGluThrTrpMetLysGluGlnGlyGluArgGluArgHisLeuThrGln 589

Qy        898 ------------AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                         |||   |||      |||::::::      ::::::      ::::::
Db        590 AspTrpGluArgLysLeuHisAlaPheGluGluGlnSerArgThrValLeuLeuGlnGlu 609
```

```
Qy        946 AAA 948
                 :::
Db        610 Arg 610

RESULT 21
US-10-321-856-81
; Sequence 81, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THER
;  FILE REFERENCE: TX-1-C2-1
;  CURRENT APPLICATION NUMBER: US/10/321,856
;  CURRENT FILING DATE:  2002-12-17
;  PRIOR APPLICATION NUMBER: 09/216,393
;  PRIOR FILING DATE: 1998-12-18
;  PRIOR APPLICATION NUMBER: 08/994,825
;  PRIOR FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 366
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-10-321-856-81

Alignment Scores:
Pred. No.:              6.82e-14      Length:        611
Score:                  325.00        Matches:       84
Percent Similarity:     54.84%        Conservative:  103
Best Local Similarity:  24.63%        Mismatches:    100
Query Match:            13.07%        Indels:        54
DB:                     4             Gaps:          11

US-09-837-344-45 (1-1482) x US-10-321-856-81 (1-611)

Qy         22 GAACAAGAGAGACGT-----------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||:::|||||||||             ||||||:::   ::::::::::::   :::
Db        292 GluArgGluArgArgArgAlaGluGluGluLysGluArgGluArgGlnGluGluGluGlu 311

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                 |||:::   |||:::   :::||||||   :::|||:::|||   :::   |||:::
Db        312 ArgGluArgArgArgValGluGluGluLysGluArgGluArgGlnGluGluGluGluArg 331

Qy        130 GAG------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
              |||      |||:::   :::||||||   ::::::::::::   |||      :::
Db        332 GluArgArgArgValGluGluGluLysAlaArgGlnArgGluGluAspGluArgGluArg 351

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT--- 240
              ||||||   :::|||:::         |||:::   :::   |||:::|||||||||
Db        352 ArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArg 371

Qy        241 ---------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                      |||   :::   :::|||::::::   :::   |||:::   |||:::
Db        372 ValGluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArgVal 391

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::||||||   ::::::::::::   :::   |||:::|||||||||      |||
Db        392 GluGluGluLysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGlu 411

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-------------- 396
                 :::   :::   :::|||   :::   |||:::||||||
Db        412 GluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGlu 431

Qy        397 ---AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
                  |||   :::   :::|||::::::   :::   |||      |||:::   :::|||
Db        432 GluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgArgValGluGluGlu 451

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
              |||   ::::::::::::   :::      :::   ||||||   :::|||:::   :::
Db        452 LysAlaArgGlnArgGluGluGluGluArgGluArgArgArgValGluGluGluLysGlu 471

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGT------GCTAAAGAAAAGTTGCAAGAA 567
```

```
                ::::::   :::   |||:::|||||||||          :::||||||   :::|||
Db        472 ArgGluArgGlnGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGlu 491

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGT------------GCTAAAGAAAAGTTGCAA 615
              :::|||   :::   |||:::|||||||||               ||||||:::   :::
Db        492 ArgGlnGluGluGluGluArgGluArgArgArgValGluGluGluLysGluArgGluArg 511

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::::::::        |||:::|||:::   :::   :::|||      :::|||:::
Db        512 GlnGluGluGluLysArgGluArgArgArgValGluGluGluLysAlaArgGlnArgGln 531

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATA 735
              :::|||   |||
Db        532 GluGluGluGly-------------------------------------------------- 535

Qy        736 GAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAG 795
                     |||   :::||||||          :::          :::
Db        536 ---------ArgGluArgGlnArgGly---------GluGluArgGluGluArgGluArg 549

Qy        796 GACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAAT 852
              :::   :::        |||:::   |||          :::   :::::::::   :::
Db        550 GluPheGlnGlnArgGluArgGluLeuLysThrArgLeuValGluLeuGlnArgGluHis 569

Qy        853 AGAGAATCTATTACAACA------AATGTTGAAGGACGAAGGGATATACAT--------- 897
                 ||||||:::   |||            ::|||   |||:::   |||
Db        570 AlaGluSerValGluThrTrpMetLysGluGlnGlyGluArgGluArgHisLeuThrGln 589

Qy        898 ------------AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                         |||   |||       |||::::::      ::::::      ::::::
Db        590 AspTrpGluArgLysLeuHisAlaPheGluGluGlnSerArgThrValLeuLeuGlnGlu 609

Qy        946 AAA 948
              :::
Db        610 Arg 610
```

RESULT 22
US-11-097-143-23688
; Sequence 23688, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23688
; LENGTH: 658
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-23688

Alignment Scores:

Pred. No.: 1.26e-13 Length: 658
Score: 321.00 Matches: 117
Percent Similarity: 47.12% Conservative: 120
Best Local Similarity: 23.26% Mismatches: 204
Query Match: 12.91% Indels: 62
DB: 6 Gaps: 17

US-09-837-344-45 (1-1482) x US-11-097-143-23688 (1-658)

```
Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              |||          |||:::|||||||||   :::   :::   ||||||:::|||
Db        186 GlnArgGlnArgGluArgGluArgArgArgGlnLeuArgArgGlnGluArgGln------ 203

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                        :::||||||:::|||:::   :::|||:::|||      |||::::::
Db        204 -----------LysLeuAlaArgGluGlnLysArgGluArgGlnArgArgLeuLysGlu 219

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
              ||||||          ::::::||||||   :::             ||||||:::||||||
Db        220 GluArg---------GlnArgLeuGlnArgGlu------------GluGlnGlnArgArg 232

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------------- 225
                 :::                |||   :::   :::
Db        233 GlnArgLeuHisHisAspGluProLysProGlnValAsnProGlnValLysProGlnVal 252

Qy        226 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
                    |||      :::|||:::   |||   |||   :::   |||:::|||
Db        253 IleAspGluLeuArgGlnArgSerGlyGluAspLeuAspLysAsnGlnThrAspGluHis 272

Qy        280 CAAGATAGACTTGCTAAAGAA---------AAGTTACAAGAGCAGCAAAGCGATTTA--- 327
              :::   :::|||       |||          ||||||:::|||:::|||
Db        273 GluGlnLysLeuArgAsnGluGlnGluLysLysLeuArgGluGluGlnGlnLysGlnArg 292

Qy        328 ---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 384
                 |||||||||::::::   :::|||:::              :::   :::   ||||||
Db        293 AspGluGlnGluGlnLysAspArgGluGluGlnAspArgLeuLysGlnGluGluGluGln 312

Qy        385 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTT 444
                 |||      ::::::::::|||:::|||   |||          :::|||
Db        313 AlaArgThrHisGlnLysGluLeuLysGluAsnGlnGluGlnGlnLeuArgGluLeuLys 332

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA--- 501
              ||||||::::::   :::::::|||:::   |||   :::|||::::::|||
Db        333 AlaLysGlnGluArgGluLysGlnGluArgAspTyrGlnGlnGlnLysArgGluHisGlu 352

Qy        502 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558
                 |||   |||::::::::::|||::::::   :::::::::        |||   |||:::
Db        353 LeuGluLeuLeuLysGlnArgGlnAlaGluAlaAspArgGlnHisAlaAlaAspGluGlu 372

Qy        559 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
                 ::::::   :::   :::      |||:::::::|||      |||      |||
Db        373 AlaGluLysLeuArgLeuGluArgIleGlnLysGlnArgGluLeuGluAlaGlnGlnArg 392

Qy        619 CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAG 678
              :::::::|||:::   :::::::||||||         ::::::      :::   ::::::
Db        393 ArgGluArgGluGluGlnArgArgLysGlnArgGluGluGlnGluGluGlnAspArgGln 412

Qy        679 GAACATGGAGATATATTAGCAGAG---------------GATTTATATGGT---CGTTTA 720
                 |||          ||||||||               ||||||||      |||:::
Db        413 AsnHisAlaLysArgLeuAlaGluGluLysArgLeuHisAspLeuTyrAlaGluArgIle 432

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
                 :::         |||         :::         :::
Db        433 ArgLeuAlaAsnThrGluArgGluLysGlnLeuAlaGluAlaHisGluAlaLysArgLeu 452

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
                    |||            :::         ::::::   :::            |||
Db        453 GluGluLeuLysLeuGlnGluGlnLeuLysLysGlnGluAspGluArgGlnGluGlnIle 472

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                 :::         |||                 |||   |||   ::::::   :::
Db        473 ArgArgGluGlnGluGluGlu--------------GluLysArgLeuGluLeuGluArg 487
```

```
Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db        488 ------LeuGluGluAlaArg-----------ArgPheGluGluLysGluLeuLysArg 501

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        502 LeuHisGluGluAsnGlnArgArgGluGluGlnLysLeuGlnArgGluArgGluIleAla 521

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        522 LeuArgGluAlaAlaGluLysLysLeuAlaGluGluGluGluMetLeuArgLysGluVal 541

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA------TATGACAATTTC 1134
Db        542 AlaGluGluGlu--------------ArgLysValLysGlnArgLeuGluAspGluMet 556

Qy       1135 CAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
Db        557 ArgGlnAlaGluGluAlaArgLysAlaLysGluAlaGluGluArgAlaAlaGluGluAla 576

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA--- 1251
Db        577 LysAlaAlaGluGlnLysArgArgValGluAlaAlaLysLysLysAlaAspGluGluVal 596

Qy       1252 ------AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
Db        597 LysAlaLysLeuGluGluLysArgArgGluTyrValThrArgIleSerAlaLeuSerPro 616

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
Db        617 GluAspGlnLysLysPheIleGluMetArgLysArgArgLysGlnLeuLysGlu---Lys 635

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA------------ 1413
Db        636 LysGluArgAspGlnArgAlaLysGluLeuLysArgIleGlnGlnAlaMetArgLeuAsp 655

Qy       1414 GACAATGAA 1422
Db        656 AspAsnGlu 658
```

```
RESULT 23
US-11-097-143-41169
; Sequence 41169, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41169
;   LENGTH: 800
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-41169

Alignment Scores:
Pred. No.:              1.26e-13      Length:          800
Score:                  321.00        Matches:         117
Percent Similarity:     47.12%        Conservative:    120
Best Local Similarity:  23.26%        Mismatches:      204
Query Match:            12.91%        Indels:          62
DB:                     6             Gaps:            17

US-09-837-344-45 (1-1482) x US-11-097-143-41169 (1-800)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
Db        328 GlnArgGlnArgGluArgGluArgArgArgGlnLeuArgArgGlnGluArgGln------ 345

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
Db        346 -----------LysLeuAlaArgGluGlnLysArgGluArgGlnArgArgLeuLysGlu 361

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
Db        362 GluArg---------GlnArgLeuGlnArgGlu-----------GluGlnGlnArgArg 374

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------------- 225
Db        375 GlnArgLeuHisHisAspGluProLysProGlnValAsnProGlnValLysProGlnVal 394

Qy        226 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db        395 IleAspGluLeuArgGlnArgSerGlyGluAspLeuAspLysAsnGlnThrAspGluHis 414

Qy        280 CAAGATAGACTTGCTAAAGAA---------AAGTTACAAGAGCAGCAAAGCGATTTA--- 327
Db        415 GluGlnLysLeuArgAsnGluGlnGluLysLysLeuArgGluGluGlnGlnLysGlnArg 434

Qy        328 ---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 384
Db        435 AspGluGlnGluGlnLysAspArgGluGluGlnAspArgLeuLysGlnGluGluGluGln 454

Qy        385 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTT 444
Db        455 AlaArgThrHisGlnLysGluLeuLysGluAsnGlnGluGlnGlnLeuArgGluLeuLys 474

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA--- 501
Db        475 AlaLysGlnGluArgGluLysGlnGluArgAspTyrGlnGlnGlnLysArgGluHisGlu 494

Qy        502 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558
Db        495 LeuGluLeuLeuLysGlnArgGlnAlaGluAlaAspArgGlnHisAlaAlaAspGluGlu 514

Qy        559 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
Db        515 AlaGluLysLeuArgLeuGluArgIleGlnLysGlnArgGluLeuGluAlaGlnGlnArg 534

Qy        619 CAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAG 678
Db        535 ArgGluArgGluGluGlnArgArgLysGlnArgGluGluGlnGluGluGlnAspArgGln 554

Qy        679 GAACATGGAGATATATTAGCAGAG--------------GATTTATATGGT---CGTTTA 720
Db        555 AsnHisAlaLysArgLeuAlaGluGluLysArgLeuHisAspLeuTyrAlaGluArgIle 574

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db        575 ArgLeuAlaAsnThrGluArgGluLysGlnLeuAlaGluAlaHisGluAlaLysArgLeu 594

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
Db        595 GluGluLeuLysLeuGlnGluGlnLeuLysLysGlnGluAspGluArgGlnGluGlnIle 614
```

```
Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db        615 ArgArgGluGlnGluGluGlu--------------GluLysArgLeuGluLeuGluArg 629

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db        630 ------LeuGluGluAlaArg------------ArgPheGluGluLysGluLeuLysArg 643

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        644 LeuHisGluGluAsnGlnArgArgGluGluGlnLysLeuGlnArgGluArgGluIleAla 663

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        664 LeuArgGluAlaAlaGluLysLysLeuAlaGluGluGluGluMetLeuArgLysGluVal 683

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAA------TATGACAATTTC 1134
Db        684 AlaGluGluGlu--------------ArgLysValLysGlnArgLeuGluAspGluMet 698

Qy       1135 CAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
Db        699 ArgGlnAlaGluGluAlaArgLysAlaLysGluAlaGluGluArgAlaAlaGluGluAla 718

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA--- 1251
Db        719 LysAlaAlaGluGlnLysArgArgValGluAlaAlaLysLysLysAlaAspGluGluVal 738

Qy       1252 ------AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
Db        739 LysAlaLysLeuGluGluLysArgArgGluTyrValThrArgIleSerAlaLeuSerPro 758

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
Db        759 GluAspGlnLysLysPheIleGluMetArgLysArgArgLysGlnLeuLysGlu---Lys 777

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA------------ 1413
Db        778 LysGluArgAspGlnArgAlaLysGluLeuLysArgIleGlnGlnAlaMetArgLeuAsp 797

Qy       1414 GACAATGAA 1422
Db        798 AspAsnGlu 800
```

RESULT 24
US-11-097-143-5154
; Sequence 5154, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5154
; LENGTH: 554
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-5154

Alignment Scores:

Pred. No.:	2.02e-13	Length:	554
Score:	318.00	Matches:	67
Percent Similarity:	62.70%	Conservative:	86
Best Local Similarity:	27.46%	Mismatches:	69
Query Match:	12.79%	Indels:	22
DB:	6	Gaps:	5

US-09-837-344-45 (1-1482) x US-11-097-143-5154 (1-554)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        291 GluGluGluLysArgLysGluGluGluArgArgArgGluGluGluArgLysArgGluGlu 310

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA------AAGTTACAAGAGCAGCAA 114
Db        311 GluArgLysArgGluGluGluArgLysArgGluGluGluArgLysArgGluGluGluArg 330

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        331 LysArgGluGluGluArgArgLysGluGluGluArgLysLysGluGluGluArgGluArg 350

Qy        175 GAACAAGAGAGA----------------------------------------CGTGCTAAA 195
Db        351 GluGluGluArgLysArgGluHisAsnArgLysLysGluGluGluArgLysArgGluGlu 370

Qy        196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
Db        371 LysArgArgLysGluGluGluLysArgLysGluGluGluArgArgLysGluGluGluArg 390

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
Db        391 LysGluGluGluArgArgLysGluGluGluArgLysGluGluGluArgArgLysGluGlu 410

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
Db        411 GluArgArgLysGluLysArgArgArgAspGluLysArgArgArgGluGluGluLysArg 430

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
Db        431 LysGluGluGluArgLysGluGluGluArgArgGluGluAlaGluArgLys---GluGlu 449

Qy        436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db        450 GluArgLysAlaGluGluArgArgLysLysGluGluArgArgArgGluGluLysArgArg 469

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
Db        470 ArgGluGluLysArgArgArgGlu------------GluGluGluArgArgLysGluGlu 485

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 615
Db        486 GluArgArgGluGluGluGluLysArgLysGluGluGluArgArgLysAsp------Glu 503

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db        504 GluArgArgArgGluGluGluLysArgLysGluGluGluArgArgGluLysGluArgArg 523

Qy        676 AAGGAACATGGA 687
Db        524 ArgGluGluGly 527
```

RESULT 25
US-10-723-860-58

```
; Sequence 58, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
;   LENGTH: 1937
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-723-860-58

Alignment Scores:
Pred. No.:              2.11e-13      Length:         1937
Score:                  317.50        Matches:        156
Percent Similarity:     42.08%        Conservative:   131
Best Local Similarity:  22.87%        Mismatches:     184
Query Match:            12.77%        Indels:         211
DB:                     5             Gaps:           35

US-09-837-344-45 (1-1482) x US-10-723-860-58 (1-1937)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||::::::   ::::::      |||      ||||||::   |||||:::::::::
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                |||:::|||::::::|||                          ::::::   |||
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||:::|||||||||      |||||||||      ||||||:::
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                  ||||||:::   |||                     :::||||||   |||
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||           ||||||||||||:::
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                     |||||||||   ||||||                     ||||||
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
              |||:::|||   |||                  :::|||   |||::::::   :::
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                                         ||||||     :::::::::  ::::::
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||         ::::::::::|||         :::::|||   ||||||
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   :::::::|||:::||||||||||   :::|||
```

```
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
                 |||:::|||:::|||                 :::   ::::|||      ::::::
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
                 :::   :::   |||:::   :::   ||||||||||:::
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540

Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                        :::|||   |||:::||||||:::::::::::::|||:::|||:::
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                                :::||||||      |||:::
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::      |||   ||||||            |||            |||:::   |||
Db       1250 LysMetCysArgSerLeuGlu------------AspGlnValSerGluLeuLysThrLys 1265

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
              :::|||           |||      ||||||               ||||||:::
Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
              |||   |||         :::   :::|||:::         :::          |||
Db       1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG------TCTATA 837
              |||   :::      :::   :::::::::      :::|||:::         :::
Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
              |||:::|||      :::   ::::::         ::::::      |||
Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
                 |||:::   :::         |||||:::::::::|||
Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy        933 ------------------------------------------------------------ 933

Db       1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy        934 --------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                           :::   |||:::   :::   |||   |||
Db       1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
                 |||   |||         :::               :::   |||   :::
Db       1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
              :::      |||   :::|||   ::::::|||   |||
Db       1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy       1087 ----------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                         |||:::   |||   |||:::|||
Db       1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------------- 1152
              :::|||:::      |||   |||
Db       1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy       1153 ----------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176
                                    :::|||:::|||            ||||||
Db       1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500
```

```
Qy       1177 TTGATAGAGAAAAATGAAAATTTA------------GATGATTTAGATGAAGGAATAGAA 1224
Db       1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy       1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
Db       1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy       1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
Db       1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy       1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
Db       1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
Db       1579 ---IleAlaGluLysAspGluGluIle------------AspGlnLeuLysArgAsnHis 1593

Qy       1459 ACTAAA 1464
Db       1594 ThrArg 1595
```

RESULT 26
US-10-723-860-106
; Sequence 106, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 106
; LENGTH: 1937
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-106

Alignment Scores:

Pred. No.:	2.11e-13	Length:	1937
Score:	317.50	Matches:	156
Percent Similarity:	42.08%	Conservative:	131
Best Local Similarity:	22.87%	Mismatches:	184
Query Match:	12.77%	Indels:	211
DB:	5	Gaps:	35

US-09-837-344-45 (1-1482) x US-10-723-860-106 (1-1937)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT-------------------------AGACTTGCTAAA 93
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG-------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540
Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
Db       1250 LysMetCysArgSerLeuGlu------------AspGlnValSerGluLeuLysThrLys 1265

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
Db       1286 GluAlaGlyGlyTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG------TCTATA 837
Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362
```

```
Qy       933 ------------------------------------------------------------ 933

Db      1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy       934 --------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                          :::   |||:::   :::   |||   |||
Db      1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy       979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
             |||   |||          :::                   :::   |||   :::
Db      1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy      1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
                :::      |||   :::|||   :::::::|||   |||
Db      1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy      1087 ------------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                           |||:::   |||   |||:::|||
Db      1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy      1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA----------------------- 1152
                :::|||:::      |||   |||
Db      1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy      1153 ------------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176
                                      :::|||:::|||           ||||||
Db      1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy      1177 TTGATAGAGAAAAATGAAAATTTA-----------GATGATTTAGATGAAGGAATAGAA 1224
             |||      :::|||:::||||||               ||||||   |||   |||
Db      1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy      1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
             :::      :::::::::   |||   ||||||||||||||      :::   ||||||
Db      1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy      1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
                :::      :::   :::   ||||||      |||      |||
Db      1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy      1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
                   :::         ::::::|||:::   |||   :::      :::
Db      1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy      1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
                |||   :::   |||   ||||||           |||:::|||      :::
Db      1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy      1459 ACTAAA 1464
             |||:::
Db      1594 ThrArg 1595

RESULT 27
US-10-028-248A-103
; Sequence 103, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
;  APPLICANT: Shimkets, Richard
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Vernet, Corine
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Malyankar, Uriel
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Li, Li
;  APPLICANT:  Taupier Jr, Raymond J
;  APPLICANT:  Kekuda, Ramesh
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Colman, Steven
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Si, Jingsheng
;  APPLICANT:  Edinger, Shlomit
;  APPLICANT:  Stone, David
;  APPLICANT:  Sciore, Paul
;  APPLICANT:  Millet, Isabelle
;  APPLICANT:  Rothenberg, Mark
;  TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 21402-222
;  CURRENT APPLICATION NUMBER: US/10/028,248A
;  CURRENT FILING DATE:  2001-12-19
;  PRIOR APPLICATION NUMBER: 60/256619
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: 60/262959
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: 60/272408
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 60/285189
;  PRIOR FILING DATE: 2001-04-20
;  PRIOR APPLICATION NUMBER: 60/308039
;  PRIOR FILING DATE: 2001-07-26
;  PRIOR APPLICATION NUMBER: 60/311266
;  PRIOR FILING DATE: 2001-08-09
;  NUMBER OF SEQ ID NOS: 211
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 103
;   LENGTH: 1961
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-028-248A-103

Alignment Scores:
Pred. No.:              2.11e-13        Length:         1961
Score:                  317.50          Matches:        145
Percent Similarity:     43.98%          Conservative:   96
Best Local Similarity:  26.46%          Mismatches:     174
Query Match:            12.77%          Indels:         133
DB:                     4               Gaps:           26

US-09-837-344-45 (1-1482) x US-10-028-248A-103 (1-1961)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::          |||            ||||||   :::||||||:::|||:::|||
Db        966 LysValThrThrGlu-----------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 -----------------TTGCAAGAACAACAA------------------------- 165
                              |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                     :::|||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                    |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                  :::||||||       |||::::::|||       :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
                 :::|||:::   |||      :::   :::|||:::||||||          :::   :::
```

```
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                                |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||      |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                    ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||              |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                 ||||||||||||||
Db       1221 --------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG------GAA 828
                  |||   |||      |||   :::            |||      |||
Db       1232 GluValLysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy        829 ATGTCTATAATAGAA--------------------------------AAAACAAATAGA 855
                 :::   |||                                  :::|||
Db       1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
                    |||   ::::::         |||         |||   |||   :::
Db       1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy        916 AAG------------------------AAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951
                                         ||||||   |||            |||   :::
Db       1292 SerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGln 1311

Qy        952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAAT 1011
              |||            :::|||            :::   :::::::::|||
Db       1312 AspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr------ 1329

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATA 1071
                 ::::::   :::   |||||||||   ::::::   :::   :::   |||
Db       1330 ------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeuGlu 1346

Qy       1072 GATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
              :::||||||||:::   :::          |||   :::            :::
Db       1347 GluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHisAlaGln 1366

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAAT 1191
                 |||                 :::   |||::::::||||||   :::
Db       1367 ValAlaAsp--------------MetLysLysLysMetGluAspSerValGlyCysLeu 1381

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT------ 1245
              |||      :::::::::          ::::::|||      |||   ||||||
Db       1382 GluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArg 1401

Qy       1246 ---GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAA 1302
                  |||||||||:::            |||   ||||||||||||            :::
Db       1402 HisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGlnGln 1421

Qy       1303 CCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                        |||      |||   |||   :::|||      :::   ::::::
Db       1422 GluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCysAsn 1439

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
                   :::   |||:::::::::||||||
Db       1440 LeuGluLysLysGlnLysLysPhe 1447
```

RESULT 28
US-10-107-782-103
; Sequence 103, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gangolli, Esha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 103
; LENGTH: 1961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-103

Alignment Scores:

Pred. No.:	2.11e-13	Length:	1961
Score:	317.50	Matches:	145
Percent Similarity:	43.98%	Conservative:	96
Best Local Similarity:	26.46%	Mismatches:	174
Query Match:	12.77%	Indels:	133
DB:	4	Gaps:	26

US-09-837-344-45 (1-1482) x US-10-107-782-103 (1-1961)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA-------------------------- 165
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG---------------------------------AGACGTGCTAAAGAAAAG 405
Db       1093 GluGluGluAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
Db       1212 LysAlaAsnLeuGluLysAlaLysGln------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
Db       1221 --------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG------GAA 828
Db       1232 GluValLysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy        829 ATGTCTATAATAGAA--------------------------------AAAACAAATAGA 855
```

```
Db       1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
Db       1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy        916 AAG------------------------AAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951
Db       1292 SerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGln 1311

Qy        952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAAT 1011
Db       1312 AspThrGlnGluLeuGluLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr------ 1329

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATA 1071
Db       1330 ------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeuGlu 1346

Qy       1072 GATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
Db       1347 GluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHisAlaGln 1366

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAAT 1191
Db       1367 ValAlaAsp--------------MetLysLysLysMetGluAspSerValGlyCysLeu 1381

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT------ 1245
Db       1382 GluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArg 1401

Qy       1246 ---GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAA 1302
Db       1402 HisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGlnGln 1421

Qy       1303 CCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
Db       1422 GluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCysAsn 1439

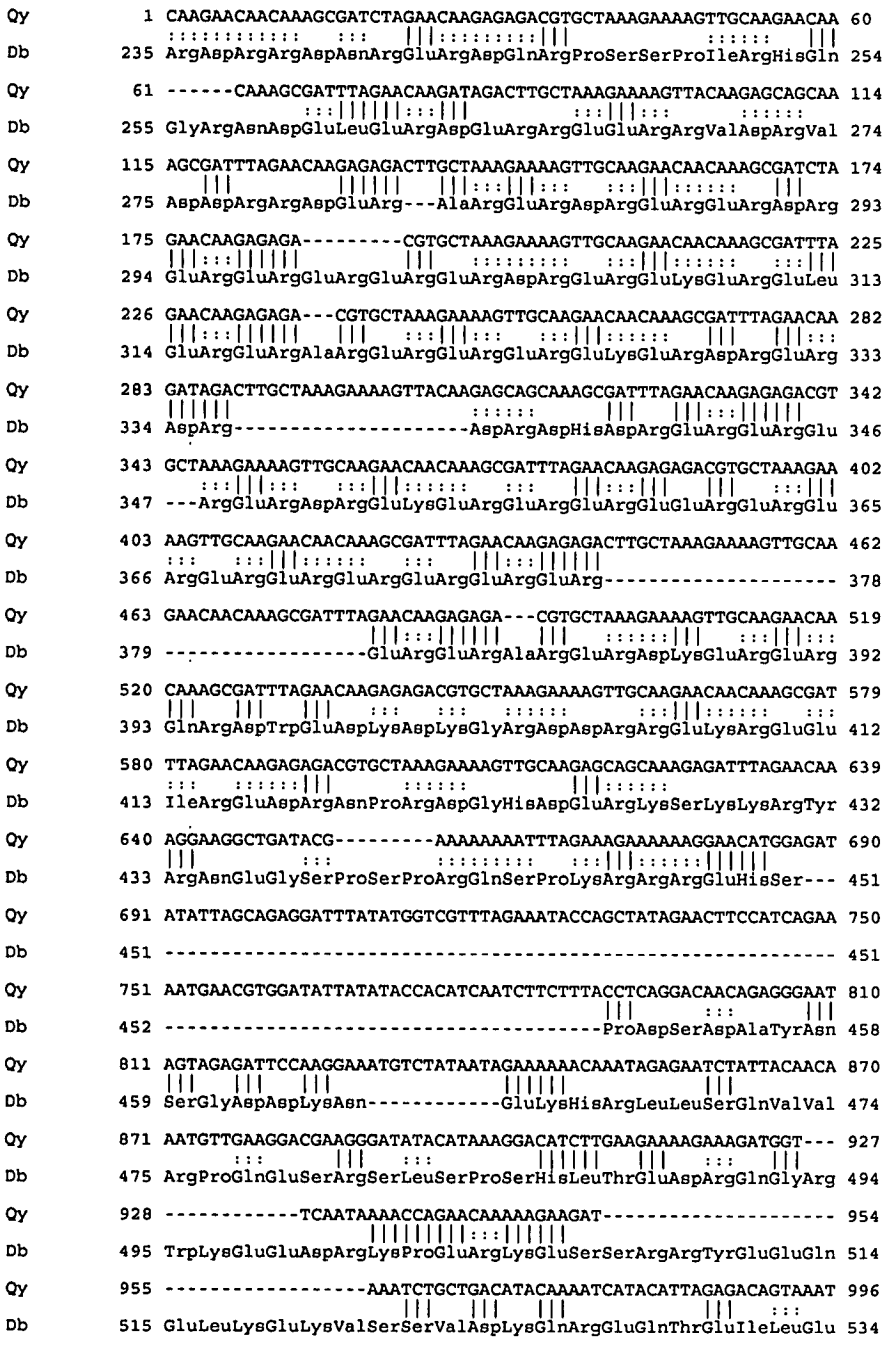
Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
Db       1440 LeuGluLysLysGlnLysLysPhe 1447
```

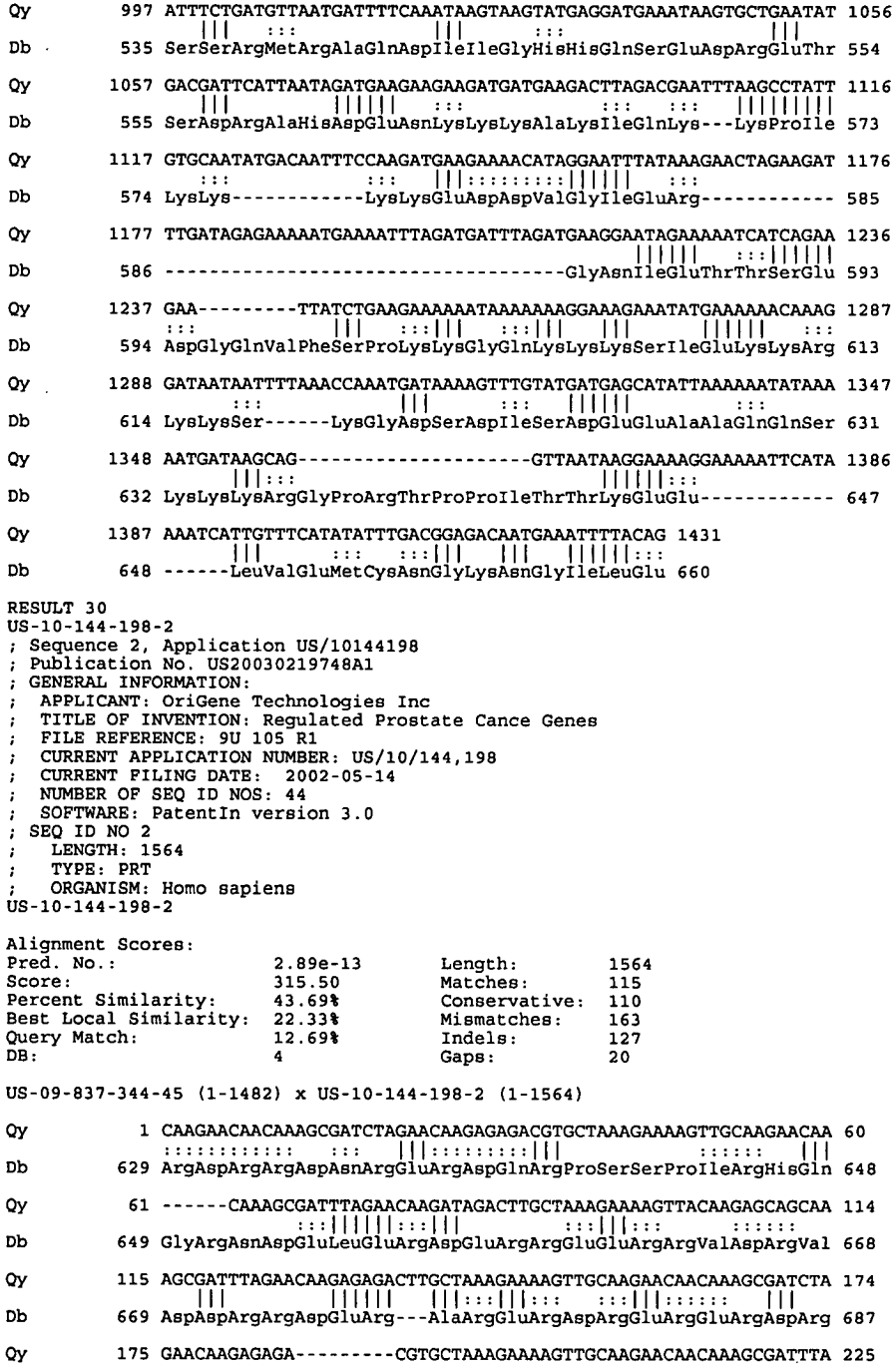
RESULT 29
US-10-341-434-95
; Sequence 95, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
; LENGTH: 1170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-95

Alignment Scores:

Pred. No.:	2.91e-13	Length:	1170
Score:	315.50	Matches:	115
Percent Similarity:	43.69%	Conservative:	110
Best Local Similarity:	22.33%	Mismatches:	163
Query Match:	12.69%	Indels:	127
DB:	4	Gaps:	20

US-09-837-344-45 (1-1482) x US-10-341-434-95 (1-1170)



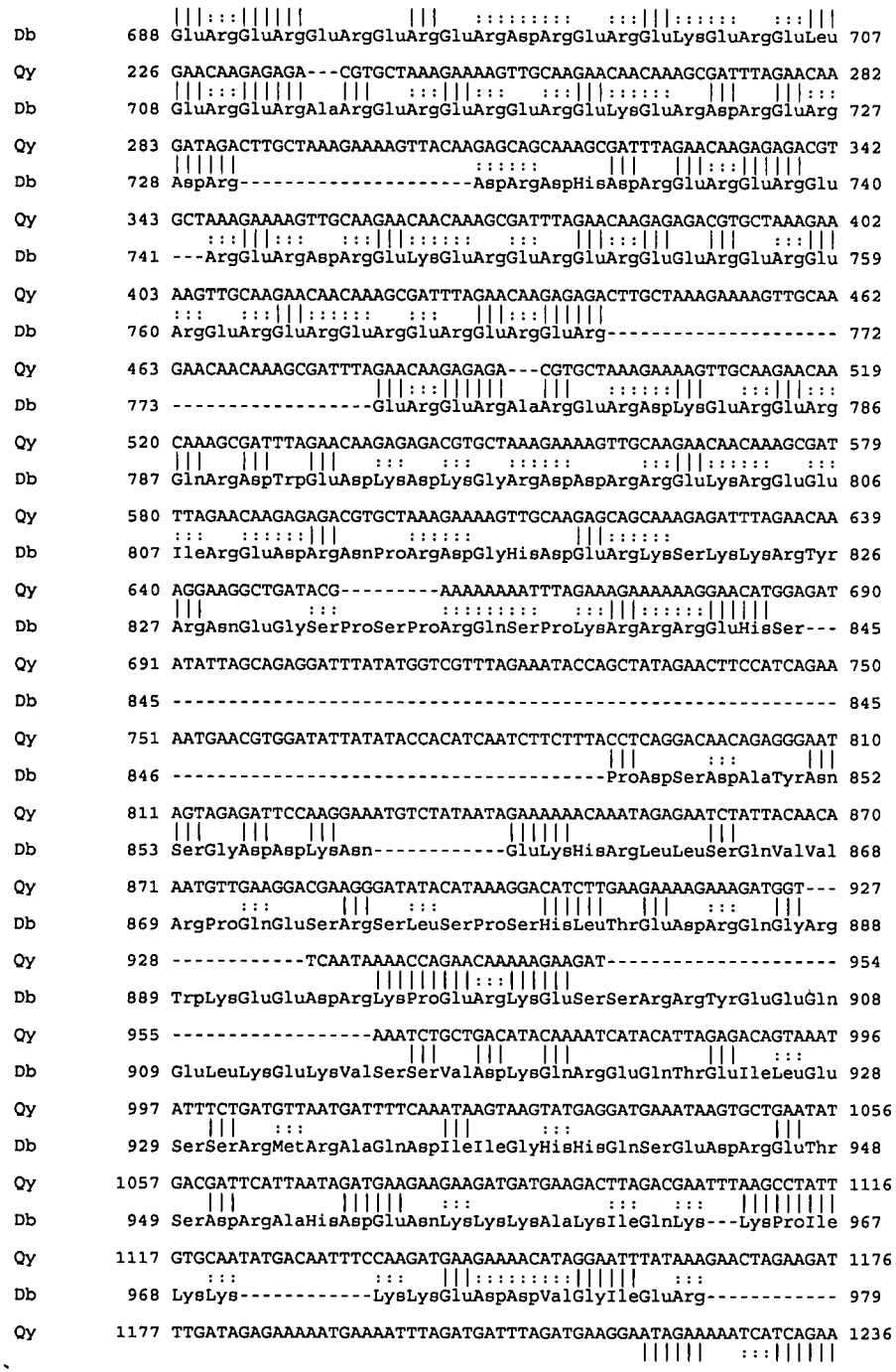


RESULT 30
US-10-144-198-2
; Sequence 2, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Alignment Scores:

Pred. No.:	2.89e-13	Length:	1564
Score:	315.50	Matches:	115
Percent Similarity:	43.69%	Conservative:	110
Best Local Similarity:	22.33%	Mismatches:	163
Query Match:	12.69%	Indels:	127
DB:	4	Gaps:	20

US-09-837-344-45 (1-1482) x US-10-144-198-2 (1-1564)



Search completed: December 21, 2005, 11:01:57
Job time : 307.352 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:38:31 ; Search time 8.43173 Seconds
(without alignments)
2507.145 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 53982 seqs, 7132107 residues

Total number of hits satisfying chosen parameters: 107964

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22867/app_query.fasta_1.3740
-DB=Published_Applications_AA_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_2_@runat_21122005_093214_22867
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	316	12.7	1976	7	US-11-069-834-54	Sequence 54, Appl
2	315	12.7	1960	7	US-11-069-834-50	Sequence 50, Appl
3	308	12.4	1976	7	US-11-069-834-52	Sequence 52, Appl
4	304	12.2	1960	7	US-11-069-834-48	Sequence 48, Appl
5	301	12.1	1652	6	US-10-995-561-663	Sequence 663, App
6	301	12.1	1938	6	US-10-995-561-661	Sequence 661, App
7	301	12.1	1938	6	US-10-995-561-662	Sequence 662, App
8	301	12.1	1954	6	US-10-995-561-660	Sequence 660, App
9	301	12.1	1972	6	US-10-995-561-664	Sequence 664, App
10	301	12.1	1972	6	US-10-995-561-666	Sequence 666, App
11	298.5	12.0	1995	7	US-11-069-834-60	Sequence 60, Appl
12	295.5	11.9	1992	7	US-11-069-834-58	Sequence 58, Appl
13	295.5	11.9	2000	7	US-11-069-834-56	Sequence 56, Appl
14	279.5	11.2	1404	6	US-10-878-556A-169	Sequence 169, App
15	275.5	11.1	2101	6	US-10-857-780-23	Sequence 23, Appl
16	266	10.7	793	7	US-11-060-914-2	Sequence 2, Appli
17	263.5	10.6	795	6	US-10-770-726-49	Sequence 49, Appl
18	254	10.2	1388	6	US-10-821-234-1143	Sequence 1143, Ap
19	251.5	10.1	718	6	US-10-878-556A-97	Sequence 97, Appl
20	251	10.1	1266	6	US-10-995-561-665	Sequence 665, App
21	241.5	9.7	885	6	US-10-793-626-1660	Sequence 1660, Ap
22	241	9.7	693	7	US-11-196-475-72	Sequence 72, Appl
23	232.5	9.3	708	7	US-11-196-475-76	Sequence 76, Appl
24	231.5	9.3	1189	7	US-11-074-176-134	Sequence 134, App
25	228	9.2	693	7	US-11-196-475-68	Sequence 68, Appl
26	228	9.2	1586	6	US-10-821-234-901	Sequence 901, App
27	223	9.0	1410	6	US-10-878-556A-136	Sequence 136, App
28	220.5	8.9	989	6	US-10-821-234-975	Sequence 975, App
29	217.5	8.7	700	7	US-11-196-475-74	Sequence 74, Appl
30	215.5	8.7	1410	6	US-10-821-234-1050	Sequence 1050, Ap
31	214.5	8.6	700	7	US-11-196-475-66	Sequence 66, Appl
32	214	8.6	663	7	US-11-196-475-70	Sequence 70, Appl
33	214	8.6	663	7	US-11-196-475-78	Sequence 78, Appl
34	211	8.5	1299	6	US-10-821-234-1145	Sequence 1145, Ap
35	208.5	8.4	898	7	US-11-099-691-7	Sequence 7, Appli
36	207.5	8.3	558	6	US-10-467-657-694	Sequence 694, App
c 37	205	7.8	3500	7	US-11-085-775-2	Sequence 2, Appli
38	201.5	8.1	691	6	US-10-131-826A-16	Sequence 16, Appl
39	201.5	8.1	703	6	US-10-821-234-963	Sequence 963, App
40	196	7.9	744	6	US-10-873-528-184	Sequence 184, App
41	195	7.8	5024	6	US-10-793-626-2964	Sequence 2964, Ap
42	193.5	7.8	664	6	US-10-878-556A-44	Sequence 44, Appl
43	192	7.7	531	7	US-11-060-914-4	Sequence 4, Appli
44	191	7.7	1095	6	US-10-793-626-3154	Sequence 3154, Ap
45	189.5	7.6	693	6	US-10-873-528-185	Sequence 185, App

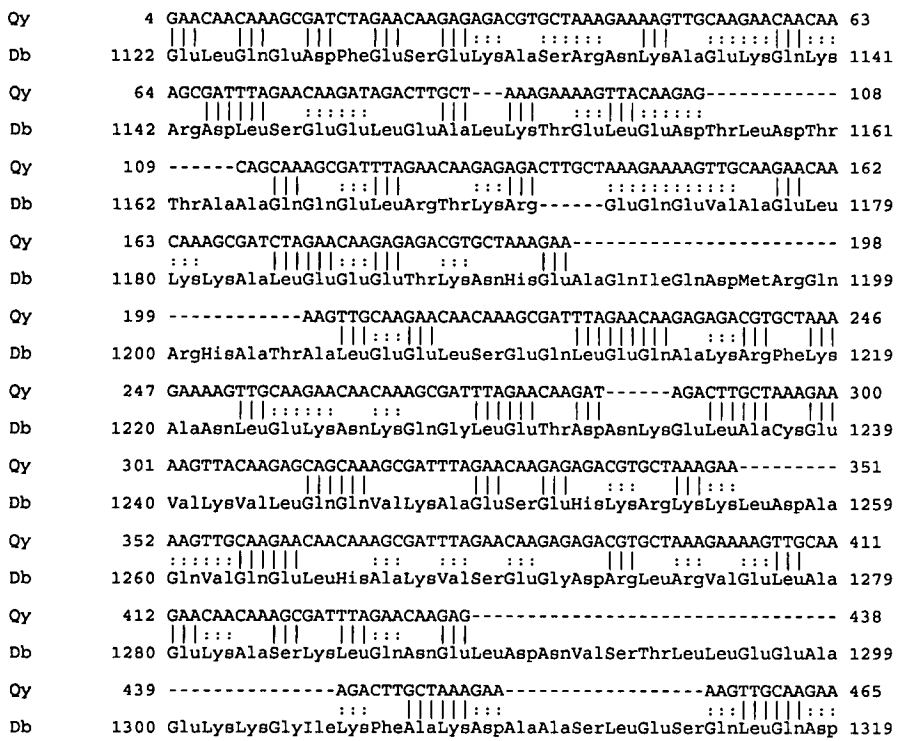
ALIGNMENTS

RESULT 1
US-11-069-834-54
; Sequence 54, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 54
; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-069-834-54

Alignment Scores:

Pred. No.:	1.4e-10	Length:	1976
Score:	316.00	Matches:	144
Percent Similarity:	40.48%	Conservative:	109
Best Local Similarity:	23.04%	Mismatches:	180
Query Match:	12.71%	Indels:	192
DB:	7	Gaps:	25

US-09-837-344-45 (1-1482) x US-11-069-834-54 (1-1976)



RESULT 2
US-11-069-834-50
; Sequence 50, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3

```
; SEQ ID NO 50
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-069-834-50

Alignment Scores:
Pred. No.:              1.58e-10      Length:          1960
Score:                  315.00        Matches:         145
Percent Similarity:     44.63%        Conservative:    100
Best Local Similarity:  26.41%        Mismatches:      168
Query Match:            12.67%        Indels:          136
DB:                     7             Gaps:            28

US-09-837-344-45 (1-1482) x US-11-069-834-50 (1-1960)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::     ||||||:::|||     |||:::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::         |||             ||||||   :::|||||||:::|||:::|||
Db        966 LysValThrThrGlu------------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
                 ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 ------------------TTGCAAGAACAACAA------------------------- 165
                               |||  |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ------------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                    :::|||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                 |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
               :::||||||      |||:::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||      :::   :::|||:::||||||          :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG---------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                               |||     :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||  |||||||||  ||||||  ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 ------------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                               ||||||   :::|||     |||||||||
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                   ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||             |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                 |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln---------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                                  ||||||||||||||      :::
Db       1221 --------------------------ThrLeuGluAsnGluArgGly---GluLeuAla 1230

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAG 825
              ::::::      :::          :::|||:::|||         |||
Db       1231 AsnGluValLysValLeuLeuGlnGlyLysGlyAspSerGluHisLysArgLysLysVal 1250

Qy        826 GAAATGTCTATAATAGAA--------------------------------AAAACAAAT 852
              |||       :::   |||                                   :::|||
Db       1251 GluAlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGlu 1270

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
                          |||   ::::::          |||          |||   |||
Db       1271 LeuAlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSer 1290

Qy        913 GAAAAG-----------------------AAAGATGGTTCAATAAAACCAGAACAAAAA 948
                  :::                      ||||||   |||           |||
Db       1291 GlnSerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeu 1310

Qy        949 GAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTT 1008
              :::|||              :::|||             :::   :::::::::|||
Db       1311 GlnAspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr--- 1329

Qy       1009 AATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068
                    ::::::   :::   ||||||||   ::::::   :::   :::   |||
Db       1330 ---------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeu 1345

Qy       1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128
                 :::|||||||||         :::|||:::      |||   |||
Db       1346 GluGluGluGluGluAlaLysHisAsnLeuGlu------LysGlnIleAlaThrLeu--- 1362

Qy       1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA 1188
                          :::         :::   |||::::::||||||   :::
Db       1363 ---------HisAlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCys 1379

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT--- 1245
                 |||      :::::::::          ::::::|||      |||   ||||||
Db       1380 LeuGluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGln 1399

Qy       1246 ------GAAGAAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                    |||||||||:::              |||   ||||||||||||
Db       1400 ArgHisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGln 1419

Qy       1300 AAACCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||       |||   |||   :::|||       :::   ::::::
Db       1420 GlnGluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCys 1437

Qy       1357 CAGGTTAATAAGGAAAAGGAAAAATTC 1383
                 :::   |||::::::::::||||||
Db       1438 AsnLeuGluLysLysGlnLysLysPhe 1446

RESULT 3
US-11-069-834-52
; Sequence 52, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
```

```
; SEQ ID NO 52
;   LENGTH: 1976
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-52

Alignment Scores:
Pred. No.:              3.61e-10      Length:         1976
Score:                  308.00        Matches:        132
Percent Similarity:     44.37%        Conservative:   136
Best Local Similarity:  21.85%        Mismatches:     168
Query Match:            12.38%        Indels:         168
DB:                     7             Gaps:           25

US-09-837-344-45 (1-1482) x US-11-069-834-52 (1-1976)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
Db        872 GluGlyGluLeuGluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIle 891

Qy         70 TTAGAACAAGAT---------------------------------------------------- 81
Db        892 LeuAlaGluGlnLeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAla 911

Qy         82 AGACTTGCT------AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA------------ 123
Db        912 ArgLeuAlaAlaLysLysGlnGluLeuGluGluIleLeuHisAspLeuGluSerArgVal 931

Qy        124 ------GAACAAGAGAGA---------CTTGCTAAAGAAAAG------------------ 150
Db        932 GluGluGluGluGluArgAsnGlnIleLeuGlnAsnGluLysLysLysMetGlnAlaHis 951

Qy        151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
Db        952 IleGlnAspLeuGluGluGlnLeuAspGluGluGluGlyAlaArgGlnLysLeuGlnLeu 971

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------------TTGCAA 258
Db        972 GluLysValThrAlaGluAlaLysIleLysLysMetGluGluGluValLeuLeuGluGlu 991

Qy        259 GAACAACAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
Db        992 AspGlnAsnSerLysPheIleLysGluLysLysLeuMetGluAspArgIleAlaGluCys 1011

Qy        316 CAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA---------------------- 348
Db       1012 SerSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAlaLysIleArgAsnLys 1031

Qy        349 --------------------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
Db       1032 GlnGluValMetIleSerAspLeuGluGluArgLeuLysLysGluGluLysThrArgGln 1051

Qy        382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
Db       1052 GluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAspLeu---GlnAspGln 1070

Qy        442 CTTGCTAAA--------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 486
Db       1071 IleAlaGluLeuGlnAlaGlnValAspGluLeuLysValGlnLeuThrLysLysGluGlu 1090

Qy        487 GAG------------------------------------------------------- 489
Db       1091 GluLeuGlnGlyAlaLeuAlaArgGlyAspAspGluThrLeuHisLysAsnAsnAlaLeu 1110

Qy        490 ---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db       1111 LysValAlaArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspPheGluSerGlu 1130

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db       1131 LysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAla 1150

Qy        601 ---AAAGAAAAGTTG----------------------------CAAGAGCAGCAAAGA 627
Db       1151 LeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThr 1170

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
Db       1171 LysArgGluGlnGluValAlaGluLeuLysLysAlaLeuGluAspGluThrLysAsnHis 1190

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
Db       1191 GluAlaGlnIleGlnAspMetArgGlnArgHisAlaThrAlaLeuGluGluLeuSerGlu 1210

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
Db       1211 GlnLeuGluGlnAlaLysArgPhe------LysAlaAsnLeuGluLysAsnLysGlnGly 1228

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCT------------------------------- 834
Db       1229 LeuGluThrAspAsnLysGluLeuAlaCysGluValLysValLeuGlnGlnValLysAla 1248

Qy        834 ------------------------------------------------------------ 834
Db       1249 GluSerGluHisLysArgLysLysLeuAspAlaGlnValGlnGluLeuHisAlaLysVal 1268

Qy        835 --------------------------ATAATAGAAAAAACAAATAGA------------ 855
Db       1269 SerGluGlyAspArgLeuArgValGluLeuAlaGluLysAlaAsnLysLeuGlnAsnGlu 1288

Qy        856 ---GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
Db       1289 LeuAspAsnValSerThrLeuLeuGlu---------GluAlaGluLysLysGlyIleLys 1305

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAA 972
Db       1306 PheAlaLysAspAlaAlaGlyLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeuGln 1325

Qy        973 AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTAT 1032
Db       1326 GluGluThrArgGlnLysLeuAsnLeuSerSer------------ArgIleArgGlnLeu 1341

Qy       1033 GAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAA 1092
Db       1342 GluGluGlu---------------LysAsnSerLeuGlnGluGlnGlnGluGluGluGlu 1356

Qy       1093 GACTTAGAC---------GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
Db       1357 GluAlaArgLysAsnLeuGluLysGlnValLeuAlaLeuGlnSerGlnLeuAlaAspThr 1376

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203
Db       1377 Lys---------------LysLysValAspAspAspLeuGlyThrIleGluSerLeuGlu 1391

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAAAAA 1254
Db       1392 GluAlaLysLysLysLeuLeuLysAspValGluAlaLeuSerGlnArgLeuGluGluLys 1411

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
Db       1412 ValLeuAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAsp 1431

Qy       1315 AGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
Db       1432 LeuThrValAspLeuAspHisGlnArgGlnIleValSerAsn------LeuGluLysLys 1449

Qy       1372 AAGGAAAAATTC 1383
Db       1450 GlnLysLysPhe 1453

RESULT 4
US-11-069-834-48
; Sequence 48, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
```

```
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 48
;   LENGTH: 1960
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-48

Alignment Scores:
Pred. No.:              5.79e-10      Length:          1960
Score:                  304.00        Matches:         149
Percent Similarity:     44.27%        Conservative:    110
Best Local Similarity:  25.47%        Mismatches:      178
Query Match:            12.22%        Indels:          148
DB:                     7             Gaps:            30

US-09-837-344-45 (1-1482) x US-11-069-834-48 (1-1960)

Qy          1 CAAGAACAA---CAAAGCGATCTAGAA---------------CAAGAGAGACGTGCT--- 39
              |||||||||   |||::::::   |||               :::|||   ||||||
Db        886 GlnGluGlnLeuGlnAlaGluThrGluLeuCysAlaGluAlaGluGluLeuArgAlaArg 905

Qy         40 ------------AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 87
                         |||::::::|||:::|||          |||||||||      |||:::
Db        906 LeuThrAlaLysLysGlnGluLeuGluGluIleCysHisAspLeuGluAla---ArgVal 924

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 147
              :::|||:::   :::   |||             |||::::::::::::   ::::::
Db        925 GluGluGluGluGluArgCysGlnTyrLeuGlnAlaGluLysLysLysMet---GlnGln 943

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207
              :::||||||   :::      ||||||:::|||     |||::::::|||||||||
Db        944 AsnIleGlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGln 963

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
                 ::::::         |||   :::   :::         ||||||:::|||   |||
Db        964 LeuGluLysValThrThrGluAlaLysLeuLys---------LysLeuGluGluAspGln 980

Qy        268 AGCGATTTAGAACAAGAT------AGACTTGCTAAAGAAAAG------------------ 303
                 :::|||           :::||||||||||||||
Db        981 IleIleMetGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuGluAspArg 1000

Qy        304 ------------------------TTACAAGAGCAGCAA--------------------- 318
                                       ::::::|||::::::
Db       1001 ValAlaGluPheThrThrAsnLeuMetGluGluGluGluLysSerLysSerLeuAlaLys 1020

Qy        319 --------------------------AGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                                         :::||||||||||:::      |||   :::|||
Db       1021 LeuLysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGlu 1040

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA--------- 402
              |||         |||:::   :::||||||:::   ||||||      |||
Db       1041 Lys---------GlnArgGlnGluLeuGluLysThrArgArgLysLeuGluGlyAspSer 1057

Qy        403 ------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
                                       :::||||||   |||   ::::::|||:::   :::
Db       1058 ThrAspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIleAlaGluLeuLysMetGln 1077

Qy        439 AGACTTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--- 489
                 |||||||||      |||:::||||||          :::   :::|||:::|||
Db       1078 ---LeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgValGluGluGluAla 1096

Qy        490 ----------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA 519
                                                                  |||      :::   :::::   |||
Db       1097 AlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluThrGlnIleSerGluLeu 1116

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              |||   |||||||||   ||||||   ::::::   |||   ::::::|||:::   |||
Db       1117 GlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAsp 1136

Qy        580 TTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG---------------------- 612
              |||   :::|||      |||   |||   :::|||
Db       1137 LeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAla 1156

Qy        613 ------CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
                   ||||||   :::         ||||||      :::      ||||||   |||
Db       1157 AlaGlnGlnGluLeuArgSerLysArgGluGlnGluValSerIleLeuLysLysThrLeu 1176

Qy        667 ----------------------------------GAAAGAAAAAAGGAACATGGAGAT 690
                                                 |||   :::::::::|||
Db       1177 GluAspGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGln 1196

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAA----------------------ATA 726
                  :::   |||:::|||      :::||||||                      :::
Db       1197 AlaVal---GluGluLeuAlaAspGlnLeuGluGlnThrLysArgValLysAlaThrLeu 1215

Qy        727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786
                    :::         |||||||||||||||                      :::
Db       1216 GluLysAlaLysGlnThrLeuGluAsnGluArgGlyGluLeuAlaAsnGluValLysAla 1235

Qy        787 TTACCTCAGGACAACAGAGGGAATAGT---------AGAGATTCCAAGGAAATGTCTATA 837
              |||   |||      :::|||:::|||         |||      |||      :::
Db       1236 LeuLeuGln---GlyLysGlyAspSerGluHisLysArgLysLysValGluAlaGlnLeu 1254

Qy        838 ATAGAA----------------------------------AAAACAAATAGAGAATCTATT 864
                  |||                                 :::|||
Db       1255 GlnGluLeuGlnValLysPheSerGluGlyGluArgValArgThrGluLeuAlaAspLys 1274

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAG------ 918
                  |||   ::::::         |||         |||   |||   :::
Db       1275 ValThrLysLeuGlnValGluLeuAspSerValThrGlyLeuLeuSerGlnSerAspSer 1294

Qy        919 ------------------AAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                                 ||||||   |||         |||   :::|||
Db       1295 LysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGln 1314

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                  :::|||         :::   :::::::::|||                  :::
Db       1315 GluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr------------Lys 1330

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              :::   :::   |||||||||   ::::::   :::   :::   |||   :::||||||
Db       1331 LeuLysGlnMetGluAspGluLysAsnSer---PheArgGluGlnLeuGluGluGluGlu 1349

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              |||         :::|||:::      |||   |||
Db       1350 GluAlaLysArgAsnLeuGlu------LysGlnIleAlaThrLeu------------His 1363

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                 :::         :::   |||::::::||||||   :::      |||
Db       1364 AlaGlnValThrAspMetLysLysLysMetGluAspGlyValGlyCysLeuGluThrAla 1383

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
              ::::::               ::::::|||      |||   ||||||         ||||||
Db       1384 GluGluAlaLysArgArgLeuGlnLysAspLeuGluGlyLeuSerGlnArgLeuGluGlu 1403

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
              |||:::            |||   ||||||||||||          :::      |||
Db       1404 LysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGlnGlnGluLeuAsp 1423

Qy       1312 AAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                    |||   |||   :::|||      :::   ::::::          :::   |||
Db       1424 AspLeuLeuValAspLeuAspHis------GlnArgGlnSerValSerAsnLeuGluLys 1441

Qy       1369 GAAAAGGAAAAATTC 1383
              :::::::::||||||
```

```
Db       1442 LysGlnLysLysPhe 1446

RESULT 5
US-10-995-561-663
; Sequence 663, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 663
;   LENGTH: 1652
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-663

Alignment Scores:
Pred. No.:              8.47e-10      Length:          1652
Score:                  301.00        Matches:         137
Percent Similarity:     42.76%        Conservative:    108
Best Local Similarity:  23.91%        Mismatches:      164
Query Match:            12.10%        Indels:          164
DB:                     6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-663 (1-1652)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::   :::|||:::   |||          ||||||
Db        646 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 665

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::      |||:::::::|||||||||
Db        666 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 685

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        686 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 705

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::     :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        706 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 725

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db        726 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 745

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db        746 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 765

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db        766 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 785

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::|||||||||::::::
Db        786 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 805

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                 ::::::   |||:::      |||   :::
Db        806 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 825

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db        826 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 845

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db        846 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 865

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db        866 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 885

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                          |||   |||:::          |||
Db        886 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 905

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db        906 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 925

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::||||||||||      |||   :::   ::::::   ||||||:::
Db        926 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 945

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
              ||||||                                 |||:::   :::|||
Db        946 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 965

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                 :::||||||
Db        966 HisLysLysLys-------------------------------LysLeuGluAlaGln 974

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                  ||||||   |||:::         |||
Db        975 ValGlnGluLeuGlnSerLysCysSerAspGly-------------------------- 985

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                     |||         |||::::::      :::|||
Db        986 --------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 996

Qy        850 AATAGA---------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
              ::::::               ||||||:::|||                ||||||:::
Db        997 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1016

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                              ::::::      ||||||   :::                |||
Db       1017 Ala-----------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1030

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
                 :::|||            :::|||      |||   :::   :::|||:::|||
Db       1031 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1050

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                        ::::::   :::   |||:::|||   ::::::         |||
Db       1051 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1065

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||   |||      ::::::|||:::              :::
Db       1066 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1082

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                 |||   |||      ::::::           |||:::|||:::|||
Db       1083 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1099

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 |||   |||::::::      :::      :::|||      |||   |||:::
Db       1100 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1119

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||           |||   ||||||||||||:::
Db       1120 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1139

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                 :::      |||      :::   |||                  |||   :::
Db       1140 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1154

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
              |||            :::   |||::::::   ||||||
```

```
Db       1155 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1167

RESULT 6
US-10-995-561-661
; Sequence 661, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: PastSEQ for Windows Version 4.0
; SEQ ID NO 661
;   LENGTH: 1938
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-661

Alignment Scores:
Pred. No.:                8.26e-10      Length:          1938
Score:                    301.00        Matches:         137
Percent Similarity:       42.76%        Conservative:    108
Best Local Similarity:    23.91%        Mismatches:      164
Query Match:              12.10%        Indels:          164
DB:                       6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-661 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::  :::|||:::  |||         ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::     |||:::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                           |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                  :::  |||  :::      |||::::::|||  ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||  |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||          ::::::  |||:::     |||  :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||  |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                         |||  |||:::          |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||  |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||  :::  ::::::  ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                          |||:::  :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                 :::||||||
Db       1252 HisLysLysLys--------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                  ||||||   |||:::      |||
Db       1261 ValGlnGluLeuGlnSerLysCysSerAspGly-------------------------- 1271

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                        |||          |||::::::      :::|||
Db       1272 --------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 1282

Qy        850 AATAGA---------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
              ::::::               ||||||:::|||                  ||||||:::
Db       1283 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1302

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                        ::::::      ||||||   :::          |||
Db       1303 Ala------------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1316

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              :::|||          :::|||      |||  :::  :::|||:::|||
Db       1317 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1336

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                    ::::::   :::   |||:::|||   ::::::         |||
Db       1337 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1351

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||   |||      ::::::|||:::          :::
Db       1352 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1368

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                 |||   |||      ::::::          |||:::|||:::|||
Db       1369 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1385

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              |||   |||::::::      :::      :::|||       |||   |||:::
Db       1386 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1405

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||                |||   ||||||||||||:::
Db       1406 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      :::   |||                        |||   :::
Db       1426 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1440

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
              |||            :::   |||::::::   ||||||
```

```
Db       1441 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1453

RESULT 7
US-10-995-561-662
; Sequence 662, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 662
;   LENGTH: 1938
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-662

Alignment Scores:
Pred. No.:              8.26e-10      Length:          1938
Score:                  301.00        Matches:         137
Percent Similarity:     42.76%        Conservative:    108
Best Local Similarity:  23.91%        Mismatches:      164
Query Match:            12.10%        Indels:          164
DB:                     6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-662 (1-1938)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::  :::|||:::  |||         ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::     |||:::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::     |||:::|||:::  |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuGluGluArgIleSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                           |||:::|||  :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                  :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||                 ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                          |||   |||:::           |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||      |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                           |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 ||||||:::                                 :::||||||
Db       1252 HisLysLysLys-------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    ||||||   |||:::        |||
Db       1261 ValGlnGluLeuGlnSerLysCysSerAspGly-------------------------- 1271

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                     |||         |||::::::      :::|||
Db       1272 --------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 1282

Qy        850 AATAGA--------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
                  ::::::           ||||||:::|||                  ||||||:::
Db       1283 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1302

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                        ::::::      ||||||   :::                  |||
Db       1303 Ala-----------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1316

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
                 :::|||            :::|||      |||   :::   :::|||:::|||
Db       1317 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1336

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                     ::::::   :::   |||:::|||   ::::::         |||
Db       1337 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1351

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   ::::||   |||      ::::::|||:::              :::
Db       1352 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1368

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                 |||   |||      ::::::           |||:::|||:::|||
Db       1369 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1385

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 |||   |||::::::      :::      :::|||      |||   |||:::
Db       1386 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1405

Qy       1246 ---------GAAGAAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||           |||   ||||||||||||:::
Db       1406 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                 :::      |||      :::   |||                        |||   :::
Db       1426 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1440

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
                 |||            :::   |||::::::   ||||||
```

```
Db       1441 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1453

RESULT 8
US-10-995-561-660
; Sequence 660, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 660
;   LENGTH: 1954
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-660

Alignment Scores:
Pred. No.:              8.25e-10      Length:          1954
Score:                  301.00        Matches:         137
Percent Similarity:     42.76%        Conservative:    108
Best Local Similarity:  23.91%        Mismatches:      164
Query Match:            12.10%        Indels:          164
DB:                     6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-660 (1-1954)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||         ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::      |||:::::::||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGlnAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAA---------------------- 195
                           |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                  :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 ----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                               :::||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                         |||          ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   |||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
              :::                            |||   |||:::              |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                     ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT-----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                           |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                  :::||||||
Db       1252 HisLysLysLys-------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                 ||||||   |||:::         |||
Db       1261 ValGlnGluLeuGlnSerLysCysSerAspGly--------------------------- 1271

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                        |||         |||::::      :::|||
Db       1272 ---------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 1282

Qy        850 AATAGA---------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
              ::::::               ||||||:::|||                  ||||||:::
Db       1283 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1302

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                          ::::::      ||||||   :::              |||
Db       1303 Ala------------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1316

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              :::|||          :::|||      |||   :::   :::|||:::|||
Db       1317 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1336

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                    ::::::   :::   |||:::|||   ::::::         |||
Db       1337 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1351

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||  |||      ::::::|||:::                :::
Db       1352 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1368

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                 |||   |||      ::::::           |||:::|||:::|||
Db       1369 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1385

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 |||   |||::::::      :::      :::|||       |||   |||:::
Db       1386 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1405

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||                ||||   ||||||||||||:::
Db       1406 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      :::   |||                      |||   :::
Db       1426 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1440

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
              |||            :::   |||::::::   ||||||
```

```
Db       1441 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1453

RESULT 9
US-10-995-561-664
; Sequence 664, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 664
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-664

Alignment Scores:
Pred. No.:              8.24e-10      Length:          1972
Score:                  301.00        Matches:         137
Percent Similarity:     42.76%        Conservative:    108
Best Local Similarity:  23.91%        Mismatches:      164
Query Match:            12.10%        Indels:          164
DB:                     6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-664 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||     |||:::  :::|||:::  |||        ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 --------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                           ||||||::::::     |||:::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::     :::|||:::||||||:::     |||:::|||:::  |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 --------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                           |||:::|||  :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                  :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 -----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                                :::||||||||:::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                        |||                 ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::           :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                             |||   |||:::          |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              |||||||||   :::|||||||||       |||   :::   ::::::   |||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                            |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                  :::||||||
Db       1252 HisLysLysLys------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                    ||||||   |||:::         |||
Db       1261 ValGlnGluLeuGlnSerLysCysSerAspGly-------------------------- 1271

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                        |||       |||::::::     :::|||
Db       1272 --------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 1282

Qy        850 AATAGA---------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
              ::::::               ||||||:::|||           ||||||:::
Db       1283 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1302

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                          ::::::     ||||||   :::           |||
Db       1303 Ala------------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1316

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              :::|||          :::|||      |||   :::   :::|||:::|||
Db       1317 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1336

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                    ::::::  :::   |||:::|||   ::::::         |||
Db       1337 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1351

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||   |||      ::::::|||:::          :::
Db       1352 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1368

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
              |||   |||      ::::::          |||:::|||:::|||
Db       1369 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1385

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 |||   |||::::::      :::      :::|||      |||   ||:::
Db       1386 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1405

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||                 |||   ||||||||||||:::
Db       1406 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
              :::      |||      :::   |||                    |||   :::
Db       1426 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1440

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
              |||             :::   |||::::::   ||||||
```

```
Db       1441 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1453

RESULT 10
US-10-995-561-666
; Sequence 666, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 666
;   LENGTH: 1972
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-666

Alignment Scores:
Pred. No.:              8.24e-10      Length:          1972
Score:                  301.00        Matches:         137
Percent Similarity:     42.76%        Conservative:    108
Best Local Similarity:  23.91%        Mismatches:      164
Query Match:            12.10%        Indels:          164
DB:                     6             Gaps:            27

US-09-837-344-45 (1-1482) x US-10-995-561-666 (1-1972)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::   |||      |||:::   :::|||:::   |||          ||||||
Db        932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy         64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
                             ||||||::::::      |||::::::|||||||||
Db        952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy        109 CAGCAA-------------------------------------------------AGCGAT 120
              ::::::                                                 |||
Db        972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspGluIleLeuValMetAsp 991

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                 :::      :::|||:::||||||:::      |||:::|||:::   |||||||||
Db        992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011

Qy        175 ---------------GAACAAGAGAGACGTGCTAAA----------------------- 195
                            |||:::|||   :::||||||
Db       1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy        196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
                    :::   |||   :::      |||::::::|||   ::::::::::::
Db       1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT----------------------- 285
              :::|||::::::   :::      ||||||   |||
Db       1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy        286 -----------------------------------AGACTTGCTAAAGAAAAG------ 303
                                                :::|||||||||::::::
Db       1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy        304 ------------TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                          |||          ::::::   |||:::      |||   :::
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspAspGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
              |||::::::|||                  |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
              |||   :::|||::::::          :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                         |||   |||:::            |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||       :::   ::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              ||||||||   :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                    ||||||                           ||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||:::                                 :::||||||
Db       1252 HisLysLysLys-------------------------------LysLeuGluAlaGln 1260

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                  ||||||   |||:::      |||
Db       1261 ValGlnGluLeuGlnSerLysCysSerAspGly-------------------------- 1271

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                                       |||          |||::::::      :::|||
Db       1272 ---------------------GluArgAlaArgAlaGluLeuAsn------AspLysVal 1282

Qy        850 AATAGA---------------GAATCTATTACA---------ACAAATGTTGAAGGACGA 885
              ::::::               ||||||:::|||                  ||||||:::
Db       1283 HisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLys 1302

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
                           ::::::      ||||||   :::                |||
Db       1303 Ala-----------------IleLysLeuAlaLysAspValAlaSerLeuSerSerGln 1316

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
              :::|||          :::|||      |||   :::   :::|||:::|||
Db       1317 LeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr 1336

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
                    ::::::   :::   |||:::|||   ::::::         |||
Db       1337 ------------LysLeuArgGlnLeuGluGluGluArgAsnSer---LeuGlnAspGln 1351

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
              |||   :::|||   |||      ::::::|||:::                :::
Db       1352 LeuAspGluGluMetGluAlaLysGlnAsnLeuGluArgHis---------IleSerThr 1368

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
                 |||   |||      ::::::          |||:::|||:::|||
Db       1369 LeuAsnIleGlnLeuSerAspSer---------LysLysLysLeuGlnAspPheAlaSer 1385

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                 |||   |||::::::      :::      :::|||      |||   |||:::
Db       1386 ThrValGluAlaLeuGluGluGlyLysLysArgPheGlnLysGluIleGluAsnLeuThr 1405

Qy       1246 ---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                      |||||||||              |||   ||||||||||||:::
Db       1406 GlnGlnTyrGluGluLysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeu 1425

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356
                 :::      |||      :::   |||                      |||   :::
Db       1426 GlnGlnGluLeuAspAspLeuValValAsp--------------LeuAspAsnGlnArg 1440

Qy       1357 CAG------------GTTAATAAGGAAAAGGAAAAATTC 1383
              |||            :::   |||::::::   ||||||
```

```
Db       1441 GlnLeuValSerAsnLeuGluLysLysGlnArgLysPhe 1453

RESULT 11
US-11-069-834-60
; Sequence 60, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 60
;   LENGTH: 1995
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-069-834-60

Alignment Scores:
Pred. No.:              1.1e-09       Length:          1995
Score:                  298.50        Matches:         150
Percent Similarity:     39.18%        Conservative:    98
Best Local Similarity:  23.70%        Mismatches:      184
Query Match:            12.00%        Indels:          201
DB:                     7             Gaps:            30

US-09-837-344-45 (1-1482) x US-11-069-834-60 (1-1995)

Qy          19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
               |||:::      |||  |||:::      |||  |||  |||  |||||||||
Db        1127 LeuLysSerLeuArgGluAlaGlnAlaAlaLeuAlaGluAlaGlnGluAspLeuGluSer 1146

Qy          79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138
               :::|||:::|||:::   |||   ::::::|||:::   ||||||   :::|||
Db        1147 GluArgValAlaArgThrLysAlaGluLysGlnArgArgAspLeuGlyGluGluLeuGlu 1166

Qy         139 GCT------------------------------------------------------- 141
               |||
Db        1167 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1186

Qy         142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
                        ::::::::::::   |||   :::      ||||||:::|||   |||
Db        1187 SerLysArgGluGlnGluValThrGluLeuLysLysThrLeuGluGluGluThrArgIle 1206

Qy         193 AAAGAA-----------------------------------AAGTTGCAAGAACAACAA 216
               |||                                     |||   |||
Db        1207 HisGluAlaAlaValGlnGluLeuArgGlnArgHisGlyGlnAlaLeuGlyGluLeuAla 1226

Qy         217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
               |||||||||   ||||||   |||        ::::::   :::          |||
Db        1227 GluGlnLeuGluGlnAlaArgArgGlyLysGlyAlaTrpGluLysThrArgLeuAlaLeu 1246

Qy         277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
               |||   :::            |||              ||||||      :::   :::
Db        1247 GluAlaGluValSerGluLeuArgAlaGluLeuSerSerLeuGlnThrAlaArgGlnGlu 1266

Qy         325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
               ||||||   ||||||       |||   :::|||||||||   |||
Db        1267 GlyGluGlnArgArgArgArgLeuGluLeuGlnLeuGlnGluValGlnGlyArgAlaGly 1286

Qy         373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
               |||   |||:::   |||      |||   ||||||||||||      |||::::::|||
Db        1287 AspGlyGluArgAlaArgAlaGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1306

Qy         430 GAACAAGAG---------------------------------AGACTTGCTAAAGAA 453
               |||                                     ||||||:::||||||
Db        1307 GluAsnValSerGlyAlaLeuAsnGluAlaGluSerLysThrIleArgLeuSerLysGlu 1326

Qy         454 -----------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
                                 :::|||   :::   |||      |||:::::|||   |||
Db        1327 LeuSerSerThrGluAlaGlnLeuHisAspAlaGlnGluLeuLeuGlnGluGluThrArg 1346

Qy         496 GCTAAAGAAAAGTTGCAAGAACAA--------------------------------CAA 522
               ||||||      |||      :::                                :::
Db        1347 AlaLysLeuAlaLeuGlySerArgValArgAlaMetGluAlaGluAlaAlaGlyLeuArg 1366

Qy         523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------------TTGCAAGAACAA 570
                     ||||||:::|||      |||:::|||:::          ||||||
Db        1367 GluGlnLeuGluGluGluAlaAlaAlaArgGluArgAlaGlyArgGluLeuGlnThrAla 1386

Qy         571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA----------------------- 606
               |||:::   |||   :::   ||||||   :::|||
Db        1387 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaGlyAlaLeuGluAla 1406

Qy         606 ------------------------------------------------------------ 606

Db        1407 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluAlaLeuThrGlnArgLeuAla 1426

Qy         607 --------------------AAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
                                   :::|||:::      :::|||   |||:::|||
Db        1427 GluLysThrGluThrValAspArgLeuGluArgGlyArgArgArgLeuGlnGlnGluLeu 1446

Qy         646 GCTGATACGAAAAAAAATTTAGAA--------------------------------AGA 672
               |||          :::||||||                                :::
Db        1447 AspAspAlaThrMetAspLeuGluGlnGlnArgGlnLeuValSerThrLeuGluLysLys 1466

Qy         673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
               :::::::::          :::|||||||||:::             |||          |||
Db        1467 GlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------Ala 1484

Qy         733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
               :::|||          |||   ||||||
Db        1485 ValGlu---------GluArgGluArgAla---------------------------- 1491

Qy         793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
                  :::      |||      |||:::::::::   :::|||:::
Db        1492 ---GluAlaGluGlyArgGluArgGluAlaArgAlaLeuSerLeuThrArgAlaLeuGlu 1510

Qy         853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
               |||                     |||   |||   ::::::
Db        1511 GluGluGln---------------GluAlaArgGluGluLeu----------------- 1519

Qy         913 GAAAAGAAAGATGGTTCAATAAAACCAGAA-----------------CAAAAAGAAGAT 954
               |||::::::::::   :::::::::   |||                |||:::|||
Db        1520 GluArgGlnAsnArgAlaLeuArgAlaGluLeuGluAlaLeuLeuSerSerLysAspAsp 1539

Qy         955 AAATCTGCTGACATACAAAATCATACATTAGAG---------ACAGTAAATATTTCTGAT 1005
                     :::      |||   ||||||              |||
Db        1540 ValGlyLysSerVal------HisGluLeuGluArgAlaCysArgValAlaGluGlnAla 1557

Qy        1006 GTTAATGATTTT------CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
               ||||||          |||::::::::   |||||||||:::::|||      |||
Db        1558 AlaAsnAspLeuArgAlaGlnValThrGluLeuGluAspGluLeuThrAlaAlaGluAsp 1577

Qy        1060 GATTCATTA-----------------------------ATAGATGAAGAAGAAGATGAT 1089
                  |||                                          |||   |||
Db        1578 AlaLysLeuArgLeuGluValThrValGlnAlaLeuLysThrGlnHisGluArgAspLeu 1597

Qy        1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
               :::      ||||||                                  :::|||
Db        1598 GlnGlyArgAspGlu--------------------------AlaGlyGluGluArgArg 1608

Qy        1150 ATAGGAATTTATAAAGAACTAGAAGATTTG---ATAGAGAAAAATGAAAATTTAGATGAT 1206
                  :::   |||:::|||   |||      :::|||::::::|||
Db        1609 ArgGlnLeuAlaLysGlnLeuArgAspAlaGluValGluArgAspGluGluArgLysGln 1628

Qy        1207 -----------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                                     :::   :::|||         |||||||||
```

```
Db       1629 ArgThrLeuAlaValAlaAlaArgLysLysLeuGluGlyGluLeuGluGluLeuLysAla 1648

Qy       1249 GAAAAAATAAAAAAAGGAAAG---AAATATGAAAAAACAAAGGAT---AATAATTTTAAA 1302
              :::          |||:::   |||   |||       |||              :::
Db       1649 GlnMetAlaSerAlaGlyGlnGlyLysGluGluAlaValLysGlnLeuArgLysMetGln 1668

Qy       1303 CCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG--- 1359
                  |||   |||:::   |||   :::::::::   :::      ::::::
Db       1669 AlaGlnMetLysGluLeuTrpArgGlu---ValGluGluThrArgThrSerArgGluGlu 1687

Qy       1360 ---------GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGAC 1410
                      |||:::|||   ||||||   :::|||   |||
Db       1688 IlePheSerGlnAsnArgGluSerGluLysArgLeuLysGlyLeu---------------- 1702

Qy       1411 GGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
                 :::   |||:::|||::::::   :::|||||||:::
Db       1703 ---GluAlaGluValLeuArgLeuGlnGluGluLeuAla 1714

RESULT 12
US-11-069-834-58
; Sequence 58, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
;  APPLICANT: CARROLL, MICHAEL C.
;  APPLICANT:  MOORE JR., FRANCIS D.
;  APPLICANT:  HECHTMAN, HERBERT B.
;  TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
;  FILE REFERENCE: CRA-002.01
;  CURRENT APPLICATION NUMBER: US/11/069,834
;  CURRENT FILING DATE:  2005-03-01
;  PRIOR APPLICATION NUMBER: 60/588,648
;  PRIOR FILING DATE: 2004-07-16
;  PRIOR APPLICATION NUMBER: 60/549,123
;  PRIOR FILING DATE: 2004-03-01
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 58
;   LENGTH: 1992
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-11-069-834-58

Alignment Scores:
Pred. No.:              1.57e-09        Length:         1992
Score:                  295.50          Matches:        151
Percent Similarity:     39.84%          Conservative:   104
Best Local Similarity:  23.59%          Mismatches:     170
Query Match:            11.88%          Indels:         215
DB:                     7               Gaps:           31

US-09-837-344-45 (1-1482) x US-11-069-834-58 (1-1992)

Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
              |||:::     |||   |||:::      |||   |||   |||   |||||||||.
Db       1123 LeuLysSerLeuArgGluAlaGlnAlaGlyLeuAlaGluAlaGlnGluAspLeuGluAla 1142

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138
              :::|||:::|||:::   |||   :::::::|||:::   ||||||   :::|||
Db       1143 GluArgValAlaArgAlaLysAlaGluLysGlnArgArgAspLeuGlyGluGluLeuGlu 1162

Qy         139 GCT------------------------------------------------------- 141
               |||
Db       1163 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1182

Qy         142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
                      :::::::::::   |||   :::      ||||||:::|||   ||||||
Db       1183 SerLysArgGluGlnGluValThrGluLeuLysLysAlaLeuGluGluGluSerArgAla 1202

Qy         193 AAAGAA----------------------------------AAGTTGCAAGAACAACAA 216
                  |||                                    |||   |||
Db       1203 HisGluValSerMetGlnGluLeuArgGlnArgHisSerGlnAlaLeuValGluMetAla 1222

Qy         217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
                 |||||||||   ||||||   |||         ::::::   :::      |||
Db       1223 GluGlnLeuGluGlnAlaArgArgGlyLysGlyValTrpGluLysThrArgLeuSerLeu 1242

Qy         277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
               |||   :::            :::                |||||      :::   :::
Db       1243 GluAlaGluValSerGluLeuLysAlaGluLeuSerSerLeuGlnThrSerArgGlnGlu 1262

Qy         325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
                  ||||||:::||||||      |||   :::|||||||||   |||         |||
Db       1263 GlyGluGlnLysArgArgArgLeuGluSerGlnLeuGlnGluValGlnGlyArgSerSer 1282

Qy         373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
               |||   |||:::   |||       |||   |||||||||||||       |||::::::|||
Db       1283 AspSerGluArgAlaArgSerGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1302

Qy         430 GAACAAGAG-----------------------------------AGACTTGCTAAAGAA 453
               |||                                         ||||||   ||||||
Db       1303 GluSerValSerThrAlaLeuSerGluAlaGluSerLysAlaIleArgLeuGlyLysGlu 1322

Qy         454 -----------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
                                :::|||   :::   |||       |||::::::|||   |||
Db       1323 LeuSerSerAlaGluSerGlnLeuHisAspThrGlnGluLeuLeuGlnGluGluThrArg 1342

Qy         496 GCTAAA---------------------------------------------------- 501
               ||||||
Db       1343 AlaLysLeuAlaLeuGlySerArgValArgAlaLeuGluAlaGluAlaAlaGlyLeuArg 1362

Qy         502 ----------------------------------------GAAAAGTTGCAAGAACAA 519
                                                      :::||||||
Db       1363 GluGlnMetGluGluGluValValAlaArgGluArgAlaGlyArgGluLeuGlnSerThr 1382

Qy         520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 555
               |||:::   |||   :::   ||||||   :::|||
Db       1383 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaAlaValLeuGluAla 1402

Qy         556 ------------------------------------------------AAGTTGCAA 564
                                                              :::|||
Db       1403 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluThrLeuThrGlnArgLeuAla 1422

Qy         565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG--- 621
               |||:::      :::|||:::     ||||||:::   :::||||||::::::
Db       1423 GluLysThrGluAlaValGluArgLeuGluArgAlaArgArgArgLeuGlnGlnGluLeu 1442

Qy         622 ---------CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
                         ||||||   |||:::|||                  ||||||:::
Db       1443 AspAspAlaThrValAspLeuGlyGlnGlnLys---GlnLeuLeuSerThrLeuGluLys 1461

Qy         673 AAAAAGGAACATGGAGAT---ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
               |||:::         |||   :::|||||||||:::            |||
Db       1462 LysGlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------ 1479

Qy         730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
               |||:::|||         :::   ||||||
Db       1480 AlaValGlu---------AspArgGluArg---------------------------- 1486

Qy         790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                  :::      |||      |||:::::::::   :::|||:::
Db       1487 ---IleGluAlaGluGlyArgGluArgGluAlaArgAlaLeuSerLeuThrArgAlaLeu 1505

Qy         850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
               |||                  |||   |||   ::::::
Db       1506 GluGluGluGln--------------GluAlaArgGluGluLeu--------------- 1515

Qy         910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAA-----------------CAAAAAGAA 951
                  |||:::::::::   :::::::::   |||                  |||:::
Db       1516 ---GluArgGlnAsnArgAlaLeuArgAlaGluLeuGluAlaLeuLeuSerSerLysAsp 1534

Qy         952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT--------------- 996
               |||         ::::::      |||   ||||||
Db       1535 AspValGlyLysAsnVal------HisGluLeuGluArgAlaArgLysAlaAlaGluGln 1552

Qy         997 ---ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
                  ||||||:::         |||:::::::::   |||||||||::::::|||
```

```
Db       1553 AlaAlaSerAspLeuArgThr---GlnValThrGluLeuGluAspGluLeuThrAlaAla 1571

Qy       1054 TATGACGATTCATTAATAGATGAA--------------------GAAGAAGATGATGAA 1092
                 |||      |||      |||                     :::   :::
Db       1572 GluAspAlaLysLeuArgLeuGluValThrValGlnAlaLeuLysAlaGlnHisGluArg 1591

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
              ||||||                          |||:::      :::|||
Db       1592 AspLeuGln-----------------GlyArgAspAspAlaGlyGluGluArgArgArg 1605

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTG---ATAGAGAAAAATGAAAATTTAGATGAT--- 1206
                 :::   |||:::|||   |||      :::|||:::::::|||
Db       1606 GlnLeuAlaLysGlnLeuArgAspAlaGluValGluArgAspGluGluArgLysGlnArg 1625

Qy       1207 --------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTA--------- 1242
                           :::   :::|||        |||||||||
Db       1626 AlaLeuAlaMetAlaAlaArgLysLysLeuGluLeuGluLeuGluGluLeuLysAlaGln 1645

Qy       1243 --------------------TCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
                                   ||||||   :::|||:::   ||||||   :::
Db       1646 ThrSerAlaAlaGlyGlnGlyLysGluGluAlaValLysGlnLeuLysLysMetGlnVal 1665

Qy       1282 ACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
                                        |||   |||:::   |||   :::::::::
Db       1666 GlnMet-----------------------LysGluLeuTrpArgGlu---ValGluGlu 1676

Qy       1342 TATAAAAATGATAAGCAG------------GTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
                 ::::::   :::               :::::::::|||   ||||||   :::|||
Db       1677 ThrArgSerSerArgAspGluMetPheThrLeuSerArgGluAsnGluLysLysLeuLys 1696

Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
                 |||                 :::   |||:::|||::::::   :::||||||:::
Db       1697 GlyLeu-----------------GluAlaGluValLeuArgLeuGlnGluGluLeuAla 1710
```

RESULT 13
US-11-069-834-56
; Sequence 56, Application US/11069834
; Publication No. US20050276811A1
; GENERAL INFORMATION:
; APPLICANT: CARROLL, MICHAEL C.
; APPLICANT: MOORE JR., FRANCIS D.
; APPLICANT: HECHTMAN, HERBERT B.
; TITLE OF INVENTION: NATURAL IGM ANTIBODIES AND INHIBITORS THEREOF
; FILE REFERENCE: CRA-002.01
; CURRENT APPLICATION NUMBER: US/11/069,834
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: 60/588,648
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: 60/549,123
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 56
; LENGTH: 2000
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-069-834-56

Alignment Scores:

Pred. No.:	1.57e-09	Length:	2000
Score:	295.50	Matches:	151
Percent Similarity:	39.84%	Conservative:	104
Best Local Similarity:	23.59%	Mismatches:	170
Query Match:	11.88%	Indels:	215
DB:	7	Gaps:	31

US-09-837-344-45 (1-1482) x US-11-069-834-56 (1-2000)

```
Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
              |||:::      |||   |||:::      |||   |||   |||   |||||||||
Db       1131 LeuLysSerLeuArgGluAlaGlnAlaGlyLeuAlaGluAlaGlnGluAspLeuGluAla 1150

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTT 138
                 :::|||:::|||:::   |||   ::::::|||:::   ||||||   :::|||
Db       1151 GluArgValAlaArgAlaLysAlaGluLysGlnArgArgAspLeuGlyGluGluLeuGlu 1170

Qy        139 GCT------------------------------------------------------- 141
                 |||
Db       1171 AlaLeuArgGlyGluLeuGluAspThrLeuAspSerThrAsnAlaGlnGlnGluLeuArg 1190

Qy        142 ---------AAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
                      ::::::::::::   |||   :::      ||||||:::|||   ||||||
Db       1191 SerLysArgGluGlnGluValThrGluLeuLysLysAlaLeuGluGluGluSerArgAla 1210

Qy        193 AAAGAA-----------------------------------AAGTTGCAAGAACAACAA 216
                 |||                                      |||   |||
Db       1211 HisGluValSerMetGlnGluLeuArgGlnArgHisSerGlnAlaLeuValGluMetAla 1230

Qy        217 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 276
                 |||||||||   ||||||   |||          ::::::   :::      |||
Db       1231 GluGlnLeuGluGlnAlaArgArgGlyLysGlyValTrpGluLysThrArgLeuSerLeu 1250

Qy        277 GAACAAGAT------------AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 324
                 |||   :::           :::                  ||||||      :::   :::
Db       1251 GluAlaGluValSerGluLeuLysAlaGluLeuSerSerLeuGlnThrSerArgGlnGlu 1270

Qy        325 TTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAACAA---------AGC 372
                 ||||||:::||||||      |||   :::|||||||||   |||         |||
Db       1271 GlyGluGlnLysArgArgArgLeuGluSerGlnLeuGlnGluValGlnGlyArgSerSer 1290

Qy        373 GATTTAGAACAAGAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
                 |||   |||:::   |||      |||   ||||||||||||      |||::::|||
Db       1291 AspSerGluArgAlaArgSerGluAlaAlaGluLysLeuGlnArgAlaGlnAlaGluLeu 1310

Qy        430 GAACAAGAG-----------------------------------AGACTTGCTAAAGAA 453
                 |||                                      ||||||   ||||||
Db       1311 GluSerValSerThrAlaLeuSerGluAlaGluSerLysAlaIleArgLeuGlyLysGlu 1330

Qy        454 ------------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
                                   :::|||   :::   |||      |||::::::|||   |||
Db       1331 LeuSerSerAlaGluSerGlnLeuHisAspThrGlnGluLeuGluGlnGluGluThrArg 1350

Qy        496 GCTAAA-------------------------------------------------------- 501
                 ||||||
Db       1351 AlaLysLeuAlaLeuGlySerArgValArgAlaLeuGluAlaGluAlaAlaGlyLeuArg 1370

Qy        502 ----------------------------------------GAAAAGTTGCAAGAACAA 519
                                                      :::||||||
Db       1371 GluGlnMetGluGluGluValValAlaArgGluArgAlaGlyArgGluLeuGlnSerThr 1390

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA---------------------- 555
                 |||:::   |||   :::   ||||||   :::|||
Db       1391 GlnAlaGlnLeuSerGluTrpArgArgArgGlnGluGluGluAlaAlaValLeuGluAla 1410

Qy        556 ------------------------------------------------AAGTTGCAA 564
                                                               :::|||
Db       1411 GlyGluGluAlaArgArgArgAlaAlaArgGluAlaGluThrLeuThrGlnArgLeuAla 1430

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG--- 621
                 |||:::         :::|||:::      |||||||:::   :::||||||::::::
Db       1431 GluLysThrGluAlaValGluArgLeuGluArgAlaArgArgArgLeuGlnGlnGluLeu 1450

Qy        622 ---------CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
                        ||||||   |||:::|||                     ||||||:::
Db       1451 AspAspAlaThrValAspLeuGlyGlnGlnLys---GlnLeuLeuSerThrLeuGluLys 1469

Qy        673 AAAAAGGAACATGGAGAT---ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                 |||:::          |||   :::||||||||:::            |||
Db       1470 LysGlnArgLysPheAspGlnLeuLeuAlaGluGluLysAlaAlaValLeuArg------ 1487

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                 |||:::|||          :::   ||||||
Db       1488 AlaValGlu---------AspArgGluArg---------------------------- 1494

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                       :::      |||      |||:::::::::   :::|||:::
```

```
Db       1495 ---IleGluAlaGluGlyArgGluArgGluAlaArgAlaLeuSerLeuThrArgAlaLeu 1513

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                  |||                 |||   |||   ::::::
Db       1514 GluGluGluGln--------------GluAlaArgGluGluLeu-------------- 1523

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAA-----------------CAAAAAGAA 951
                 |||:::::::::   :::::::::   |||                 |||:::
Db       1524 ---GluArgGlnAsnArgAlaLeuArgAlaGluLeuGluAlaLeuLeuSerSerLysAsp 1542

Qy        952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT-------------- 996
                 |||         ::::::      |||   ||||||
Db       1543 AspValGlyLysAsnVal------HisGluLeuGluArgAlaArgLysAlaAlaGluGln 1560

Qy        997 ---ATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
                  ||||||:::         |||:::::::::   |||||||||::::::|||
Db       1561 AlaAlaSerAspLeuArgThr---GlnValThrGluLeuGluAspGluLeuThrAlaAla 1579

Qy       1054 TATGACGATTCATTAATAGATGAA--------------------GAAGAAGATGATGAA 1092
                 |||     |||     |||                     :::   :::
Db       1580 GluAspAlaLysLeuArgLeuGluValThrValGlnAlaLeuLysAlaGlnHisGluArg 1599

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
                 ||||||                    |||:::    :::|||
Db       1600 AspLeuGln------------------GlyArgAspAspAlaGlyGluGluArgArgArg 1613

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTG---ATAGAGAAAAATGAAAATTTAGATGAT--- 1206
                  :::   |||:::|||   |||      :::|||::::::|||
Db       1614 GlnLeuAlaLysGlnLeuArgAspAlaGluValGluArgAspGluGluArgLysGlnArg 1633

Qy       1207 ---------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTA--------- 1242
                            :::   :::|||         |||||||||
Db       1634 AlaLeuAlaMetAlaAlaArgLysLysLeuGluLeuGluLeuGluLeuLysAlaGln 1653

Qy       1243 --------------------TCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
                                   ||||||   :::|||:::   ||||||   :::
Db       1654 ThrSerAlaAlaGlyGlnGlyLysGluGluAlaValLysGlnLeuLysLysMetGlnVal 1673

Qy       1282 ACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
                                        |||   |||:::   |||   :::::::::
Db       1674 GlnMet------------------------LysGluLeuTrpArgGlu---ValGluGlu 1684

Qy       1342 TATAAAAATGATAAGCAG------------GTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
                 ::::::   :::                :::::::::|||   ||||||   :::|||
Db       1685 ThrArgSerSerArgAspGluMetPheThrLeuSerArgGluAsnGluLysLysLeuLys 1704

Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
                 |||                 :::   |||:::|||::::::   :::|||||:::
Db       1705 GlyLeu-----------------GluAlaGluValLeuArgLeuGlnGluGluLeuAla 1718

RESULT 14
US-10-878-556A-169
; Sequence 169, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 169
;   LENGTH: 1404
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: humangp/chr12-q14221
;   DATABASE ENTRY DATE: 2003-04-22
US-10-878-556A-169

Alignment Scores:
Pred. No.:              1.1e-08        Length:        1404
```

```
Score:                  279.50         Matches:       115
Percent Similarity:     45.49%         Conservative:  142
Best Local Similarity:  20.35%         Mismatches:    207
Query Match:            11.24%         Indels:        101
DB:                     6              Gaps:          21

US-09-837-344-45 (1-1482) x US-10-878-556A-169 (1-1404)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::          ||||||:::|||:::|||:::            |||:::   :::
Db        740 GluGlnAlaLeuGlnAspLeuGlnGlnGlnArgGlnLeuAsnThrAspLeuGluLeuArg 759

Qy         61 CAAAGCGATTTA---GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 117
                ::::::|||   :::|||       :::   ||||||   :::         :::
Db        760 AlaThrGluLeuSerLysGlnLeuGluMetGluLysGluIleValSerSerThrArgLeu 779

Qy        118 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
              ||||||:::::::::       |||   |||   ::::::::::::   :::       |||
Db        780 AspLeuGlnLysLysSerGluAlaLeuGluSerIleLysGlnLysLeuThrLysGlnGlu 799

Qy        178 CAAGAGAGACGT----------------------------GCTAAAGAAAAGTTGCAA 207
              :::|||::::::                                :::   |||:::|||
Db        800 GluGluLysLysIleLeuLysGlnAspPheGluThrLeuSerGlnGluThrLysIleGln 819

Qy        208 GAACAACAA--------------------------AGCGATTTAGAACAAGAGAGACGT 240
              ::::::                          ::::::|||::::::   :::
Db        820 HisGluGluLeuAsnAsnArgIleGlnThrThrValThrGluLeuGlnLysValLysMet 839

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA--- 297
              ||||||   |||       :::   |||   :::   :::|||:::|||:::|||
Db        840 GluLysGluAlaLeuMetThrGluLeuSerThrVal---LysAspLysLeuSerLysVal 858

Qy        298 ---GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG----------------- 336
                 :::   |||:::      :::|||:::   |||:::|||
Db        859 SerAspSerLeuLysAsnSerLysSerGluPheGluLysGluAsnGlnLysGlyLysAla 878

Qy        337 --------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA--- 366
                                        :::     |||   :::||||||   |||
Db        879 AlaIleLeuAspLeuGluLysThrCysLysGluLeuLysHisGlnLeuGlnValGlnMet 898

Qy        367 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 426
              ::::::   |||::::::::::::      |||:::   |||::::::::::::
Db        899 GluAsnThrLeuLysGluGlnLysGluLeuLysLysSerLeuGluLysGluLysGluAla 918

Qy        427 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA------------CAAAGC 474
                 |||   :::|||             :::|||||||||             |||:::
Db        919 SerHisGlnLeuLysLeuGluLeuAsnSerMetGlnGluGlnLeuIleGlnAlaGlnAsn 938

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                 |||:::|||      :::   :::::::::||||||         ::::::|||:::
Db        939 ThrLeuLysGlnAsnGluLysGluGluGlnGlnLeuGlnGlyAsnIleAsnGluLeuLys 958

Qy        535 CAAGAGAGACGTGCTAAAGAAAAG---TTGCAAGAACAACAAAGCGATTTAGAA------ 585
                 |||           |||:::|||   ::::::      |||   :::|||:::
Db        959 GlnSerSerGluGlnLysLysLysGlnIleGluAlaLeuGlnGlyGluLeuLysIleAla 978

Qy        586 ------CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                    |||:::         :::   |||||||||:::|||                |||
Db        979 ValLeuGlnLysThrGluLeuGluAsnLysLeuGlnGlnGlnLeuThrGlnAlaAlaGln 998

Qy        640 AGGAAGGCTGATACGAAAAAA--------------------AATTTAGAAAGAAAAAAG 678
                 |||      |||:::                      |||   |||:::   :::
Db        999 GluLeuAlaAlaGluLysGluLysIleSerValLeuGlnAsnAsnTyrGluLysSerGln 1018

Qy        679 GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAA 738
              |||           |||      |||   |||||||||                  |||
Db       1019 GluThrPheLysGlnLeuGlnSerAspPheTyrGlyArg------------GluSerGlu 1034

Qy        739 CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGAC 798
              |||   :::      :::            :::      :::   ||||||   |||:::
Db       1035 LeuLeuAlaThrArgGlnAspLeuLysSerValGluGluLysLeuSerLeuAlaGlnGlu 1054

Qy        799 AACAGAGGGAATAGTAGAGAT------TCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
```

```
Db       1055 AspLeuIleSerAsnArgAsnGlnIleGlyAsnGlnAsnLysLeuIleGlnGluLeuLys 1074

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912

Db       1075 ThrAlaLysAlaThrLeuGluGlnAspSerAlaLys---------LysGluGlnGlnLeu 1091

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGAT------AAATCTGCTGAC 966

Db       1092 GlnGluArgCysLysAlaLeuGlnAspIleGlnLysGluLysSerLeuLysGluLysGlu 1111

Qy        967 ATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGT 1026

Db       1112 LeuValAsnGluLysSerLysLeuAlaGluIleGluGluIleLys------------Cys 1127

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086

Db       1128 ArgGlnGluLysGluIleThrLys-----------------LeuAsnGluGluLeuLys 1141

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146

Db       1142 SerHisLysLeuGluSerIleLysGluIleThr---------AsnLeuLysAspAlaLys 1158

Qy       1147 AACATAGGAATTTATAAAGAACTAGAA----------------------------GAT 1176

Db       1159 GlnLeuLeuIleGlnGlnLysLeuGluLeuGlnGlyLysAlaAspSerLeuLysAlaAla 1178

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236

Db       1179 ValGluGlnGluLysArgAsnGlnGlnIleLeuLysAspGlnValLysLysGluGluGlu 1198

Qy       1237 GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAA---------AAAACAAAG 1287

Db       1199 GluLeuLysLysGluPheIleGluLysGluAlaLysLeuHisSerGluIleLysGluLys 1218

Qy       1288 GATAATAATTTTAAACCAAATGATAAA--------------------------AGTTTG 1320

Db       1219 GluValGlyMetLysLysHisGluGluAsnGluAlaLysLeuThrMetGlnIleThrAla 1238

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Db       1239 LeuAsnGluAsnLeuGlyThrValLysLysGluTrpGlnSerSerGlnArgArgValSer 1258

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Db       1259 GluLeuGluLysGlnThrAspAspLeuArgGlyGluIleAlaValLeuGluAlaThrVal 1278

Qy       1441 GAGTTATCTGAAGAT 1455

Db       1279 GlnAsnAsnGlnAsp 1283

RESULT 15
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
;  APPLICANT: ROTH, RICHARD B.
;  APPLICANT:  BRAUN, ANDREAS
;  APPLICANT:  KAMMERER, STEFAN M.
;  APPLICANT:  NELSON, MATTHEW ROBERTS
;  APPLICANT:  RENELAND, RIKARD HENRY
;  APPLICANT:  HOYAL-WRIGHTSON, CAROLYN R.
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: SEQ-4069-CP
;  CURRENT APPLICATION NUMBER: US/10/857,780
;  CURRENT FILING DATE:  2004-05-28
;  PRIOR APPLICATION NUMBER: 10/723,681
;  PRIOR FILING DATE: 2003-11-25
;  PRIOR APPLICATION NUMBER: 60/490,234
;  PRIOR FILING DATE: 2003-07-24
;  PRIOR APPLICATION NUMBER: 60/525,239
;  PRIOR FILING DATE: 2003-11-25
;  NUMBER OF SEQ ID NOS: 4962
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
;   LENGTH: 2101
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-857-780-23

Alignment Scores:
Pred. No.:                1.65e-08      Length:          2101
Score:                    275.50        Matches:         133
Percent Similarity:       42.94%        Conservative:    86
Best Local Similarity:    26.08%        Mismatches:      168
Query Match:              11.08%        Indels:          123
DB:                       6             Gaps:            24

US-09-837-344-45 (1-1482) x US-10-857-780-23 (1-2101)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69

Db        472 GlnSerSerIleSerAsnLeuSerGlnAlaLysGluGluLeu---GluGlnAlaSerGln 490

Qy         70 TTAGAACAAGATAGACTTGCTAAAGAA------------------------------- 96

Db        491 AlaHisGlyAlaArgLeuThrAlaGlnValAlaSerLeuThrSerGluLeuThrThrLeu 510

Qy         97 ------AAGTTACAAGAGCAGCAAAGCGATTTAGAA---CAAGAGAGACTTGCTAAAGAA 147

Db        511 AsnAlaThrIleGlnGlnGlnAspGlnGluLeuAlaGlyLeuLysGlnGlnAlaLysGlu 530

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207

Db        531 LysGlnAlaGlnLeuAlaGlnThrLeuGlnGlnGlnGluGlnAlaSerGlnGlyLeuArg 550

Qy        208 GAACAA------------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255

Db        551 HisGlnValGluGlnLeuSerSerSerLeuLysGlnLysGluGlnGlnLeuLysGluVal 570

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGA------------------CTTGCTAAA 297

Db        571 AlaGluLysGlnGluAlaThrArgGlnAspHisAlaGlnGlnLeuAlaThrAlaAlaGlu 590

Qy        298 GAAAAG---------TTACAAGAG------------------CAGCAAAGCGATTTAGAA 330

Db        591 GluArgGluAlaSerLeuArgGluArgAspAlaAlaLeuLysGlnLeuGluAlaLeuGlu 610

Qy        331 CAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAACAA-------------------- 366

Db        611 LysGluLysAlaAlaLysLeuGluIleLeuGlnGlnGlnLeuGlnValAlaAsnGluAla 630

Qy        367 ------------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAA------------GAA 402

Db        631 ArgAspSerAlaGlnThrSerValThrGlnAlaGlnArgGluLysAlaGluLeuSerArg 650

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 462

Db        651 LysValGluGluLeuGlnAlaCysValGluThrAlaArgGlnGluGlnHisGluAlaGln 670

Qy        463 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 522

Db        671 AlaGlnValAlaGluLeuGluLeuGlnLeuArgSer--------------GluGlnGln 685

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAAGAACAACAAAGC 576

Db        686 LysAlaThrGluLysGluArgValAlaGlnGluLysAspGlnLeuGlnGluGlnLeuGln 705

Qy        577 GATTTA----------------------------------GAACAAGAGAGACGTGCT 600

Db        706 AlaLeuLysGluSerLeuLysValThrLysGlySerLeuGluGluGluLysArgArgAla 725

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Db        726 AlaAspAlaLeuGluGluGlnGlnArgCysIleSerGluLeuLysAlaGluThrArgSer 745
```

```
Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--- 717
              :::|||:::   |||          |||   |||:::      ||||||
Db        746 LeuValGluGlnHisLysArgGluArgLysGluLeuGluGluGluArgAlaGlyArgLys 765

Qy        718 ---TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                  ||||||      :::::::|||
Db        766 GlyLeuGluAlaArgLeuLeuGlnLeuGlyGlu----------------------Ala 777

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCT 834
              ||||||:::           :::              |||:::     |||
Db        778 HisGlnAla-----------GluThrGluValLeuArgArgGluLeuAlaGluAlaMet 793

Qy        835 ATAATAGAAAAAACAAATAGA---------GAATCTATTACAACAAATGTTGAAGGACGA 885
                  :::   |||              |||  :::      |||
Db        794 AlaAlaGlnHisThrAlaGluSerGluCysGluGlnLeuValLysGluValAlaAlaTrp 813

Qy        886 AGGGATATACATAAA---GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAA 942
              ||||||   ::::::          ||||||   :::   |||::::::      |||
Db        814 ArgAspGlyTyrGluAspSerGlnGlnGluGluAlaGlnTyrGlyAlaMetPheGlnGlu 833

Qy        943 CAAAAA--------------------GAAGATAAATCTGCTGACATACAAAATCATACA 981
              |||                     |||          :::::::|||
Db        834 GlnLeuMetThrLeuLysGluGluCysGluLysAlaArgGlnGluLeuGlnGluAlaLys 853

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT--- 1038
              :::      |||          ::::::   |||||||||:::   ::::::
Db        854 GluLysValAlaGlyIleGluSerHisSerGluLeuGlnIleSerArgGlnGlnAsnLys 873

Qy       1039 ------GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT--- 1089
                 |||:::   |||          :::|||        :::|||   :::
Db        874 LeuAlaGluLeuHisAlaAsnLeuAlaArgAlaLeuGlnGlnValGlnGluLysGluVal 893

Qy       1090 -----------------GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
                           :::||||||        :::
Db         894 ArgAlaGlnLysLeuAlaAspAspLeuSerThrLeuGln--------------------- 906

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA---------CTAGAAGATTTGATA 1182
                 |||:::          ||||||       ||||||   |||:::
Db         907 ---------GluLysMetAlaAlaThrSerLysGluValAlaArgLeuGluThrLeuVal 923

Qy       1183 GAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTA 1242
                |||                     |||      |||   :::|||   ||||||
Db         924 ArgLys-----------------AlaGlyGluGlnGlnGluThrAlaSerArgGluLeu 937

Qy       1243 TCTGAAGAAAAAATAAAAAAAGGAAAGAAA 1272
                :::|||      :::   |||   :::
Db         938 ValLysGluProAlaArgAlaGlyAspArg 947
```

RESULT 16
US-11-060-914-2
; Sequence 2, Application US/11060914
; Publication No. US20050261184A1
; GENERAL INFORMATION:
; APPLICANT: Kaufman, Paul L
; APPLICANT: Geiger, Benjamin
; APPLICANT: Bershadsky, Alexander
; APPLICANT: Borras, Teresa
; TITLE OF INVENTION: Method for Treating Glaucoma
; FILE REFERENCE: 960296.00128
; CURRENT APPLICATION NUMBER: US/11/060,914
; CURRENT FILING DATE: 2005-02-18
; PRIOR APPLICATION NUMBER: US 60/545,722
; PRIOR FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: US 60/545,723
; PRIOR FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-060-914-2

Alignment Scores:

Pred. No.:	5.89e-08	Length:	793
Score:	266.00	Matches:	120
Percent Similarity:	43.55%	Conservative:	103
Best Local Similarity:	23.44%	Mismatches:	199
Query Match:	10.70%	Indels:	90
DB:	7	Gaps:	20

US-09-837-344-45 (1-1482) x US-11-060-914-2 (1-793)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG---------- 48
              :::   ::::::   |||   |||:::|||         :::||||||
Db        182 GluAsnLysLysGluAspLysGluLysGluGluGluGluGluLysProLysArgGly 201

Qy         49 ---TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
                 :::   |||   |||   ::::::          :::   :::||||||
Db        202 SerIleGlyGluAsnGlnValGluVal-----------MetValGluGluLysThrThr 217

Qy        106 GAGCAGCAAAGCGAT----------------------------TTAGAACAAGAGAGA 135
              |||   |||   :::                           :::      |||
Db        218 GluSerGlnGluGluThrValValMetSerLeuLysAsnGlyGlnIleSerSerGluGlu 237

Qy        136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------------CTAGAACAA 180
                  :::|||:::   :::||||||   ||||||                 :::
Db        238 ProLysGlnGluGluGluArgGluGlnGlySerAspGluIleSerHisHisGluLysMet 257

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||       ||||||:::  :::  :::::::::   ||||||   |||   |||
Db        258 GluGluGluAspLysGluArgAlaGluAlaGluArgAlaArgLeuGluAlaGluGluArg 277

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
                    |||:::::::::   :::|||                |||      |||:::
Db        278 ------GluArgIleLysAlaGluGlnAspLysLysIleAlaAspGluArgAlaArgIle 295

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              :::   :::|||:::   :::      |||:::|||||||||      |||
Db        296 GluAlaGluGluLysAlaAlaAlaGlnGluArgGluArgArgGluAlaGlu--------- 312

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||::::::       |||:::|||:::||||||   |||        |||:::|||
Db        313 GluArgGluArgMetArgGluGluGluLysArgAlaAlaGlu---------GluArgGln 329

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                    |||:::|||:::   |||   |||          |||:::|||
Db        330 ArgIleLysGluGluGluLysArgAlaAlaGlu---------GluArgGlnArgIleLys 346

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::|||:::||||||   |||          |||:::|||          |||:::|||
Db        347 GluGluGluLysArgAlaAlaGlu---------GluArgGlnArgIleLysGluGluGlu 363

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              :::||||||   |||      :::   |||::::::      |||:::|||   :::|||
Db        364 LysArgAlaAlaGlu------GluArgGlnArgAlaArgAlaGluGluGluGluLysAla 381

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              |||   :::   |||::::::   :::   ||||||::::::          |||
Db        382 LysValGluGluGlnLysArgAsnLysGlnLeuGluGluLys-----------LysHis 397

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT------ 714
              ::::::      |||      |||:::   :::   :::   :::   |||
Db        398 AlaMetGlnGluThrLysIleLysGlyGluLysValGluGlnLysIleGluGlyLysTrp 417

Qy        715 --------------------------CGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
                                       :::|||:::      ::::::
Db        418 ValAsnGluLysLysAlaGlnGluAspLysLeuGlnThrAlaValLeuLysLysGlnGly 437

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
              |||   |||:::|||      :::                |||      |||
Db        438 Glu---GluLysGlyThrLysVal--------------------GlnAlaLysArgGlu 449

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
              ::::::   |||          |||:::        |||   |||
```

```
Db        450 LysLeuGlnGluAspLysProThrPheLysLysGluGluIleLysAspGluLysIleLys 469

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927

Db        470 LysAspLysGluProLysGluGluVal---LysSerPheMetAspArgLysLysGlyPhe 488

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA------ 981

Db        489 ThrGluValLysSerGlnAsnGlyGluPheMetThrHisLysLeuLysHisThrGluAsn 508

Qy        982 ----------------------------TTAGAGACAGTAAATATTTCTGATGTTAAT 1011

Db        509 ThrPheSerArgProGlyGlyArgAlaSerValAspThrLysGluAlaGluGlyAlaPro 528

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATA----------------------AGT 1047

Db        529 GlnValGluAlaGlyLysArgLeuGluGluLeuArgArgArgArgGlyGluThrGluSer 548

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107

Db        549 GluGluPheGluLysLeuLysGlnLysGlnGlnGluAlaAlaLeuGluLeuGluGluLeu 568

Qy       1108 AAGCCTATTGTGCAATATGACAAT------TTCCAAGATGAAGAAAACATAGGAATTTAT 1161

Db        569 LysLysLysArgGluGluArgArgLysValLeuGluGluGluGluGlnArgArgLysGln 588

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221

Db        589 GluGluAlaAspArgLysLeuArgGluGluGluGluLysArgArgLeuLysGluGluIle 608

Qy       1222 GAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA---------AAAGGAAAGAAA 1272

Db        609 GluArgArgArgAlaGluAlaAlaGluLysArgGlnLysMetProGluAspGlyLeuSer 628

Qy       1273 TATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332

Db        629 AspAspLysLysProPheLysCysPheThrProLysGlySerSerLeuLysIleGluGlu 648

Qy       1333 ATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368

Db        649 ArgAlaGluPheLeuAsn---LysSerValGlnLys 659

RESULT 17
US-10-770-726-49
; Sequence 49, Application US/10770726
; Publication No. US20050266409A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Brown, Eugene
;  APPLICANT:  Liu, Wei
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION:  CANCERS
;  FILE REFERENCE: AM101079 (031896-010000)
;  CURRENT APPLICATION NUMBER: US/10/770,726
;  CURRENT FILING DATE:  2004-02-04
;  NUMBER OF SEQ ID NOS: 48640
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
;   LENGTH: 795
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-770-726-49

Alignment Scores:
Pred. No.:              7.9e-08       Length:          795
Score:                  263.50        Matches:         105
Percent Similarity:     42.47%        Conservative:    112
Best Local Similarity:  20.55%        Mismatches:      177
Query Match:            10.60%        Indels:          117
DB:                     6             Gaps:            17

US-09-837-344-45 (1-1482) x US-10-770-726-49 (1-795)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db         18 GlnGluLysLysArgArgLysGluGlnGluGluLysAlaGluIleLysArgLeuLysAsn 37

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Db         38 SerAspAspArgAspSerLysArg---------AspSerLeuGluGlu------GlyGlu 52

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Db         53 LeuArgAspHisArgMet---------GluIleThrIleArgAsnSerProTyrArgArg 69

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------GATTTAGAACAA 231

Db         70 GluAspSerMetGluAspArgGlyGluGluAspAspSerLeuAlaIleLysProProGln 89

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291

Db         90 GlnMetSerArgLysGluLysValHisHisArg--------------LysAspGluLys 104

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351

Db        105 ArgLysGluLysArgArgHisArgSerHisSerAlaGluGlyGlyLysHisAlaArgVal 124

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411

Db        125 LysGluLysGluArg-----------GluHisGluArgArgLysArgHisArgGluGlu 140

Qy        412 GAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 471

Db        141 GlnAspLysAlaArgArgGluTrpGluArgGlnLysArgArgGluMetAlaArgGluHis 160

Qy        472 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531

Db        161 Ser---------------ArgArgGluArgAspArgLeuGluGlnLeuGluArgLysArg 175

Qy        532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591

Db        176 GluArgGluArgLysMetArgGluGlnGlnLysGluGlnArgGluGlnLysGluArgGlu 195

Qy        592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGAT 651

Db        196 ArgArgAlaGluGluArgArgLysGluArgGlu---------------------------- 206

Qy        652 ACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711

Db        206 ------------------------------------------------------------ 206

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771

Db        207 AlaArgArgGluValSerAla------------------------------------- 213

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG 831

Db        214 ---HisHisArgThrMetArgGluAspTyrSerAspLysValLysAlaSerHisTrpSer 232

Qy        832 TCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGAT 891

Db        233 ArgSerProProArgProProArgGluArgPheGluLeuGly---AspGlyArgLysPro 251

Qy        892 ATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951

Db        252 GlyGluAlaArgProAlaArgAlaGlnLysProAlaGlnLeuLysGluGluLysMetGlu 271

Qy        952 GATAAATCT------GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005

Db        272 GluArgAspLeuLeuSerAspLeuGln----------------------------Asp 281

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065

Db        282 IleSerAspSerGluArgLysThrSerSerAlaGluSerSerSerAlaGluSerGlySer 301

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATAT 1125
```

```
Db        302 GlySerGluGluGluGluGluGluGluGluGluGluGluGlySerThrSerGlu 321

Qy       1126 GACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAG 1185
Db        322 GluSerGluGluGluGluGlu----------------------GluGluGluGluGlu 333

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
Db        334 GluGluGluThrGlySerAsnSerGluGluAlaSerGluGlnSerAlaGluGluValSer 353

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
Db        354 GluGluGluMetSerGluAspGlu------GluArgGluAsnGluAsnHisLeuLeuVal 371

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
Db        372 ValProGluSerArgPheAsp--------------------ArgAspSerGlyGluSer 384

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
Db        385 GluGluAlaGluGlu-----------------GluValGlyGluGlyThrProGlnSer 398

Qy       1426 TTACAGATCGTGGAT----------------------------------------GAGTTA 1446
Db        399 SerAlaLeuThrGluGlyAspTyrValProAspSerProAlaLeuSerProIleGluLeu 418

Qy       1447 TCTGAAGATATAACTAAATATTTTATGAAACTA 1479
Db        419 LysGlnGluLeuProLysTyrLeuProAlaLeu 429

RESULT 18
US-10-821-234-1143
; Sequence 1143, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1143
;   LENGTH: 1388
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1143

Alignment Scores:
Pred. No.:              2.22e-07      Length:          1388
Score:                  254.00        Matches:         112
Percent Similarity:     42.75%        Conservative:    106
Best Local Similarity:  21.96%        Mismatches:      144
Query Match:            10.21%        Indels:          148
DB:                     6             Gaps:            21

US-09-837-344-45 (1-1482) x US-10-821-234-1143 (1-1388)

Qy           1 CAAGAACAACAAAGCGATCTAGAA------CAAGAGAGACGTGCTAAAGAAAAGTTGCAA 54
Db         582 GluGluArgLysGluArgLeuGluSerLeuAsnIleGlnArgGluLysGluGluLeuGlu 601

Qy          55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCT---AAAGAAAAGTTACAAGAGCAG 111
Db         602 GlnArgGluAlaGluLeuGlnLysValArgLysAlaGluGluGluArgLeuArgGlnGlu 621

Qy         112 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA--------------- 156

Db         622 AlaLysGluArgGluLysGluArgIleLeuGlnGluHisGluGlnIleLysLysLysThr 641

Qy         156 ------------------------------------------------------------ 156

Db         642 ValArgGluArgLeuGluGlnIleLysLysThrGluLeuGlyAlaLysAlaPheLysAsp 661

Qy         157 ----------------------------------------------GAACAACAAAGC 168

Db         662 IleAspIleGluAspLeuGluGluLeuAspProAspPheIleMetAlaLysGlnValGlu 681

Qy         169 GATCTAGAACAAGAGAGACGT---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db         682 GlnLeuGluLysGluLysLysGluLeuGlnGluArgLeuLysAsnGlnGluLysLysIle 701

Qy         226 GAACAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db         702 AspTyrPheGluArgAlaLysArgLeuGluGluIleProLeuIleLysSerAlaTyrGlu 721

Qy         280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT--------------- 324
Db         722 GluGlnArgIleLysAspMetAspLeuTrpGluGlnGlnGluGluGluArgIleThrThr 741

Qy         325 ----------------------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db         742 MetGlnLeuGluArgGluLysAlaLeuGluHisLysAsnArgMetSerArgMetLeuGlu 761

Qy         361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db         762 AspArgAspLeuPheValMetArgLeuLysAlaAlaArgGlnSerValTyrGluGluLys 781

Qy         421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAA---------AAGTTGCAAGAACAACAA 471
Db         782 LeuLysGlnPheGluGluArgLeuAlaGluGluArgHisAsnArgLeuGluGluArgLys 801

Qy         472 AGCGATTTAGAACAAGAGAGACGTGCT---------AAAGAAAAGTTGCAAGAACAACAA 522
Db         802 ArgGlnArgLysGluGluArgArgIleThrTyrTyrArgGluLysGluGluGluGluGln 821

Qy         523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
Db         822 -----------------ArgArgAlaGluGluGlnMetLeuLysGluArgGluGluArg 835

Qy         583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
Db         836 GluArgAlaGluArgAlaLysArg------GluGluGluLeuArgGluTyrGlnGluArg 853

Qy         643 AAGGCTGATACGAAAAAA------AATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696
Db         854 ---------ValLysLysLeuGluGluValGluArgLysLysArgGlnArgGlu------ 868

Qy         697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756
Db         869 --------------------LeuGluIleGluGluArgGluArgArgArgGluGluGlu 881

Qy         757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
Db         882 Arg------ArgLeuGlyAspSerSerLeuSerArgLysAspSerArgTrpGlyAspArg 899

Qy         817 GATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT 876
Db         900 AspSer-----------------GluGlyThrTrpArgLysGlyProGluAlaAspSer 913

Qy         877 GAAGGACGAAGG--------------GATATACATAAAGGACATCTTGAAGAAAAGAAA 921
Db         914 GluTrpArgArgGlyProProGluLysGluTrpArgArgGlyGluGlyArgAspGluAsp 933

Qy         922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
Db         934 ArgSerHisArgArgAspGluGluArgProArgArgLeuGlyAspAspGluAspArg--- 952

Qy         982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
```

```
Db        953 ----------------------------------------------------------Glu 953

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGAC------ 1095
Db        954 ProSerLeuArgProAspAspAspArgValProArgArgGlyMetAspAspAspArgGly 973

Qy       1096 ----------------------------------------TTAGACGAATTTAAGCCT 1113
Db        974 ProArgArgGlyProGluGluAspArgPheSerArgArgGlyAlaAspAspAspArgPro 993

Qy       1114 ATTGTGCAATATGACAATTTCCAAGAT----------------------GAAGAAAAC 1149
Db        994 ------SerTrpArgAsnThrAspAspAspArgProProArgArgIleAlaAspGluAsp 1011

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
Db       1012 ArgGlyAsnTrpArgHisAlaAspAspAspArgProProArgArgGlyLeuAspGlu--- 1030

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAA 1239
Db       1031 AspArgGlySerTrpArgThrAlaAspGlu 1040

RESULT 19
US-10-878-556A-97
; Sequence 97, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
;   LENGTH: 718
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: humangp/chr7-q9unh0
;   DATABASE ENTRY DATE: 2003-01-07
US-10-878-556A-97

Alignment Scores:
Pred. No.:              3.3e-07       Length:         718
Score:                  251.50        Matches:        116
Percent Similarity:     45.27%        Conservative:   104
Best Local Similarity:  23.87%        Mismatches:     175
Query Match:            10.11%        Indels:         91
DB:                     6             Gaps:           23

US-09-837-344-45 (1-1482) x US-10-878-556A-97 (1-718)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db         44 GlnTyrGlnGlnSerMetGlnLeuGluGluArgAlaGluGlnIleArgSerLysSerHis 63

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db         64 LeuIleGlnValGluArgGluLysMetGlnMetGlu---LeuSerHisLysArgAlaArg 82

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db         83 ValGluLeuGluArgAlaAlaSerThrSerAlaArgAsnTyrGluArgGluValAspArg 102

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        103 AsnGlnGluLeuLeuThrArgIleArgGlnLeuGln---------GluArgGluAlaGly 119

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA------------ 288
Db        120 AlaGluGluLysMetGlnGluGln---------LeuGluArgAsnArgGlnCysGlnGln 136

Qy        289 ---------CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG--- 336
Db        137 AsnLeuAspAlaAlaSerLysArgLeuArgGluLysGluAspSerLeuAlaGlnAlaGly 156

Qy        337 ---------AGACGTGCTAAAGAAAAGTTGCAAGAA---CAACAAAGCGATTTAGAACAA 384
Db        157 GluThrIleAsnAlaLeuLysGlyArgIleSerGluLeuGlnTrpSerValMetAspGln 176

Qy        385 GAGAGACGTGCT-----------------AAAGAAAAGTTGCAAGAACAACAAAGCGAT 426
Db        177 GluMetArgValLysArgLeuGluSerGluLysGlnGluLeuGlnGlu---GlnLeuAsp 195

Qy        427 TTAGAACAAGAGAGACTT------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        196 LeuGlnHisLysLysCysGlnGluAlaAsnGlnLysIleGlnGluLeuGlnAla------ 213

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        214 SerGlnGluAlaArgAlaAspHis------GluGlnGlnIleLysAspLeuGluGln--- 230

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db        231 ---------LysLeuSerLeuGlnGluGlnAspAlaAlaIleValLysAsnMetLysSer 247

Qy        601 AAA---GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA 657
Db        248 GluLeuValArgLeuProArgLeuGluArgGluLeuLysGlnLeuArgGluGluSer--- 266

Qy        658 AAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT--- 714
Db        267 AlaHisLeuArgGluMetArgGluThrAsnGlyLeuLeuGlnGluGluLeuGluGlyLeu 286

Qy        715 ------------------------CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
Db        287 GlnArgLysLeuGlyArgGlnGluLysMetGlnGluThrLeuValGlyLeuGluLeuGlu 306

Qy        751 AATGAACGTGGATAT---TATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
Db        307 AsnGluArgLeuLeuAlaLysLeuGlnSerTrpGluArgLeuAspGlnThrMetGlyLeu 326

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCT------ATAATAGAAAAAACAAATAGAGAA--- 858
Db        327 SerIleArgThrProGluAspLeuSerArgPheValValGluLeuGlnGlnArgGluLeu 346

Qy        859 ---------------------TCTATTACAACAAATGTTGAAGGA--------------- 882
Db        347 AlaLeuLysAspLysAsnSerAlaValThrSerSerAlaArgGlyLeuGluLysAlaArg 366

Qy        883 ------------------CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db        367 GlnGlnLeuGlnGluGluLeuArgGlnValSerGlyGlnLeuLeuGluGluArgLys--- 385

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
Db        386 ---------LysArgGluThrHisGluAlaLeuAlaArgArgLeuGlnLysArgValLeu 402

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
Db        403 LeuLeuThrLysGluArgAspGlyMetArgAlaIleLeuGlySerTyrAspSerGluLeu 422

Qy       1045 AGT---GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
Db        423 ThrProAlaGluTyrSerProGlnLeuThrArgArgMetArgGluAlaGluAspMet--- 441

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        442 ------------ValGlnLysValHisSerHisSerAlaGluMetGluAlaGlnLeuSer 457

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221
Db        458 GlnAlaLeuGluGluLeu-----------------------------------GlyGly 465

Qy       1222 GAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
```

```
Db        466 GlnLysGlnArgAlaAspMetLeuGluMetGluLeuLysMetLeuLysSerGlnSerSer 485

Qy       1282 ACAAAGGATAATAATTTT 1299
Db        486 SerAlaGluGlnSerPhe 491

RESULT 20
US-10-995-561-665
; Sequence 665, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 665
;   LENGTH: 1266
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-665

Alignment Scores:
Pred. No.:                3.21e-07        Length:          1266
Score:                    251.00          Matches:         114
Percent Similarity:       42.18%          Conservative:    91
Best Local Similarity:    23.46%          Mismatches:      175
Query Match:              10.09%          Indels:          106
DB:                       6               Gaps:            19

US-09-837-344-45 (1-1482) x US-10-995-561-665 (1-1266)

Qy          19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
Db         844 LeuLeuGlnValThrArgGlnGluGluGluMetGlnAlaLysGluAspGluLeuGlnLys 863

Qy          79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG------ 132
Db         864 ThrLysGluArgGlnGlnLysAlaGluAsnGluLeuLysGluLeuGluGlnLysHisSer 883

Qy         133 AGACTTGCTAAAGAAAAG------TTGCAAGAACAA---CAAAGCGATCTAGAA------ 177
Db         884 GlnLeuThrGluGluLysAsnLeuLeuGlnGluGlnLeuGlnAlaGluThrGluLeuTyr 903

Qy         178 ----------------------------------CAAGAGAGACGTGCTAAAGAAAAG 201
Db         904 AlaGluAlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGluGluIle 923

Qy         202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
Db         924 LeuHisGluMetGluAlaArgLeuGluGluGluGluAspArgGlyGlnGlnLeuGlnAla 943

Qy         262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db         944 GluArgLysLysMetAlaGlnGlnMetLeu---------AspLeuGluGluGln------ 958

Qy         322 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
Db         959 ---LeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGlu 977

Qy         382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA---------AGCGATTTAGAA 432
Db         978 -----------AlaLysIleLysLysLeuGluAspGluIleLeuValMetAspAspGln 993

Qy         433 CAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATTTA------ 480
Db         994 AsnAsnLysLeuSerLysGluArgLysLeuGluGluArgIleSerAspLeuThrThr 1013

Qy         481 ---------GAACAAGAGAGACGTGCTAAAGAA------AAGTTGCAAGAACAACAAAGC 525
Db        1014 AsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLysHisGlu 1033

Qy         526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---CAAAGCGATTTA 582
Db        1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGlnGluLeu 1053

Qy         583 GAACAAGAGAGACGTGCTAAAGAA------------AAGTTGCAAGAGCAGCAAAGAGAT 630
Db        1054 GluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIleAlaAsp 1073

Qy         631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
Db        1074 LeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGlu--------- 1090

Qy         691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
Db        1091 ------------------------GluLeuGlnAlaAlaLeuAlaArgLeuAspAspGlu 1102

Qy         751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
Db        1102 ------------------------------------------------------------ 1102

Qy         811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
Db        1103 ------------------------IleAlaGlnLysAsnAsnAlaLeuLysLysIleArg 1114

Qy         871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930
Db        1115 GluLeuGluGly------------HisIleSerAspLeuGlnGluAspLeuAspSerGlu 1130

Qy         931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
Db        1131 ArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAla 1150

Qy         991 GTA------AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
Db        1151 LeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAla 1170

Qy        1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
Db        1171 LysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHis 1190

Qy        1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
Db        1191 GluAlaGlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGlu 1210

Qy        1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
Db        1211 GlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysAsnLysGlnThrLeuGlu 1230

Qy        1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
Db        1231 LysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluVal 1250

Qy        1285 AAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATAT 1344
Db        1251 ------------------------------------------GluHis---LysLysLys 1255

Qy        1345 AAAAATGATAAGCAGGTT 1362
Db        1256 LysLeuGluAlaGlnVal 1261

RESULT 21
US-10-793-626-1660
; Sequence 1660, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
```

```
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
;   LENGTH: 885
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-1660

Alignment Scores:
Pred. No.:              1.04e-06      Length:         885
Score:                  241.50        Matches:        99
Percent Similarity:     39.53%        Conservative:   101
Best Local Similarity:  19.57%        Mismatches:     147
Query Match:            9.71%         Indels:         159
DB:                     6             Gaps:           17

US-09-837-344-45 (1-1482) x US-10-793-626-1660 (1-885)

Qy          7 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 66
              |||:::         |||   ||||||:::   |||||:::      :::   :::
Db        489 GlnArgGlnThrAlaGluPheGluArgGlnPheLysGluGlnLysThrGlnAlaGluGln 508

Qy         67 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 126
                |||||         |||   ::::::          :::      :::
Db        509 LeuSerGluGlnTyrPheSerAlaSerGlnGlnTyrAsnAsnLeuLysGluGlnValHis 528

Qy        127 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGA 186
                  |||      ::::::|||:::   |||::::::   |||:::   |||
Db        529 HisHisGluLeuGluLeuAspArgLeuLysThrGlnGluAlaHisLeuLysAsnGluHis 548

Qy        187 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 246
                |||:::   :::   :::::::::|||      |||      :::::::|||
Db        549 ---------GluGluPheGluPheGluLysAsnAspGlyTyrGlnSerAspLysSerLys 565

Qy        247 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTA 306
              |||   |||:::|||:::|||:::   |||   :::          :::::::::|||
Db        566 GluThrLeuLysGluLysGlnAsnHisLeuIleGlu---------IleGlnGlnGlnLeu 582

Qy        307 CAAGAGCAGCAAAGCGATTTAGAACAA---------------------------------- 333
              ::::::   :::||||||:::|||:::
Db        583 LysGlnLeuGluSerAspIleGluArgTyrThrGlnLeuSerLysGluGlyLysAlaSer 602

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---------------- 378
                  :::   :::::::::|||   ::::::||||||||||||
Db        603 ThrHisGlnThrGlnGlnGlnLeuHisGlnLysGlnSerAspLeuAlaValValLysGlu 622

Qy        379 ------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
                    :::   :::::::::      |||:::|||   :::|||   ||||||   |||
Db        623 ArgIleLysSerGlnLysGlnValTyrGluArgLeuAspLysGlnLeuSerAspSerGlu 642

Qy        433 CAAGAGAGACTT---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
              ::::::::::::         ||||||::::::         ||||||
Db        643 ArgGlnLysIleGluValAsnGluLysIleLysLeuPheAsnSerAspGluMetMetGly 662

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
              :::   |||   ||||||||||:::||||||         ::::::|||:::
Db        663 LysAspAlaPheGluLysLeuLysGluGln---------IleGlnGlnGlnGluAsnVal 679

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
              ::::::   |||   :::|||   |||:::::::::|||:::|||:::
Db        680 ArgGlnAsnLeuAsnGlnGlnLeuSerGluIleLysGlnGlnArgLys------------ 695

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669

Db        695 ------------------------------------------------------------ 695

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729

                                                  ||||||      ::::::||||||
Db        696 ---------------------------------AspLeuAsnGluLysIleGluIle--- 703

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                              ||||||                ||||||      |||
Db        704 --------------------AsnGluSerGlnLeuGlnLysCysHisGlnAspIleLeu 716

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                  :::|||
Db        717 SerIleGluAsn---------------------------------------------- 720

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                                                                    |||
Db        721 ------------------------------------------------------HisTyr 722

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              ::::::   |||                          |||:::   |||         :::
Db        723 GlnAspIleLys--------------------AlaLysGlnSerLysLeuAspValLeu 735

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
                |||||   ::::::   :::|||          :::   :::|||::::::
Db        736 IleAsnHisAlaIleAspHisLeuAsn-----------AspThrTyrGlnLeuThr--- 750

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                  :::              |||||||||                          |||
Db        751 ---ValGluArgAlaArgMetGluTyrAsp----------------------SerAsp 761

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA--- 1146
              |||   :::|||      :::      |||:::                    :::|||
Db        762 GluThrIleAspAsnLeuArgLysLysValLysLeuThrLysMetThrIleAspGluLeu 781

Qy       1147 ---------AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA--------- 1188
                      |||:::         ::::::   |||:::|||   |||:::
Db        782 GlyProValAsnLeuAsnAlaIleGluGlnPheGluGluLeuAsnGluArgTyrThrPhe 801

Qy       1189 ---AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                |||||         ||||||   |||      |||         :::   :::
Db        802 LeuAsnGluGlnArgThrAspLeuArgGluAlaLysGluThrLeuGluGlnIleIleHis 821

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              |||                          :::|||      :::      ||||||
Db        822 Glu-----------------------MetAspLysGluValGluGlyArgPheLysThr 833

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
                    ::::::   |||               :::|||
Db        834 ThrPheHisAlaValGlnAspHisPheThrThrValPheLys----------------- 847

Qy       1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
                                                  :::|||   |||
Db        848 ---------------------------------GlnLeuPheGlyGlyGlyGlnAlaGlu 856

Qy       1426 TTACAGATCGTGGATGAG 1443
                |||::::::   ::::::
Db        857 LeuArgLeuThrGluAsp 862

RESULT 22
US-11-196-475-72
; Sequence 72, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
```

```
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-72

Alignment Scores:
Pred. No.:                1.15e-06      Length:          693
Score:                    241.00        Matches:         96
Percent Similarity:       44.52%        Conservative:    111
Best Local Similarity:    20.65%        Mismatches:      157
Query Match:              9.69%         Indels:          101
DB:                       7             Gaps:            16

US-09-837-344-45 (1-1482) x US-11-196-475-72 (1-693)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
Db        230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy         73 GAACAAGATAGACTTGCTAAA----------------GAAAAGTTACAAGAGCAGCAA 114
Db        247 GluLeuAspAsnValHisGluSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        267 AspGlnLeuGlu---LysAlaThrAspGluGluHisArgLysGluIleGluSerGlnVal 285

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
Db        286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy        235 AGACGTGCTAAAGAAAAGTTG------CAAGAACAACAAAGCGATTTAGAACAAGATAGA 288
Db        304 ---LysAlaGlnGlnLysLeuAspPheSerGluAspAsnLeuAspIleGlnArgAspThr 322

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
Db        323 Val---ArgGluLysIleGlnGluAspIleAsnGluIleAsnLysGluLysAsnLeuPro 341

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
Db        342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
Db        362 GluAspLeuGlnGluGlnLeu--------------------------------LysGlu 370

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
Db        371 ThrSerAspGluAsnGlnLysArgGluIleGluLysGlnIleGluIleLysLysSerAsp 390

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db        391 GluGluLeuLeuLysSerLysAspProLysAlaLeuAspLeuAsnArgAspLeuAsnSer 410

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db        411 LysAlaSerSerLysGluLysIleLysGlyLysGluLysGluIleValLysGluLysSer 430

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
Db        431 ----------------------LysValSerLeuGlyAspLeu--------------- 437

Qy        709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
Db        438 ------------------------------------AspAsnAspGluThrLeuMet 444

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC---AAG 825
Db        445 ThrProGluAspGlnLysLeuSerGluAspLysLysLeuAspSerLysLysAsnLeuLys 464

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGA 885
Db        465 ProValSerGluIleGluArgValAsnGluIleSerLysSerAsnAsnAsnGlu------ 482

Qy        886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAA 945
Db        483 ---------------------------------------------ValSerLysSerSer 487

Qy        946 AAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGAT 1005
Db        488 ProLeuAspLysProSerTyrSerAspIleAspSerLysGluValValAspAsnLysAsp 507

Qy       1006 GTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCA 1065
Db        508 ValAsnLeuGlnGluThrLysProGlnAlaLysSerGlnSerThrSerLeuAsnGlnAsp 527

Qy       1066 TTAATAGATGAAGAAGAAGATGATGAAGAC---------TTAGACGAATTTAAGCCTATT 1116
Db        528 LeuIleThrMetSerIleAspSerSerAsnProValPheLeuGluValIleAspProIle 547

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGAT 1176
Db        548 Thr--------------------------AsnLeuGlyMetLeuGln------------ 554

Qy       1177 TTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAA 1236
Db        555 --------------------LeuIleAspLeuAsnThrGlyValArg------LeuLys 565

Qy       1237 GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
Db        566 GluSerThrGlnGlnGlyIleGlnArgTyrGlyIleTyrGluArgGluLysAsp-LeuVa 585

Qy       1297 TTTAAACCAAATG------------ATAAAAGTTTGTATGATGAGCATATTAAAAAATAT 1344
Db        585 lValIleLysMetAspSerGlyLysAlaLysLeuGlnIleLeuAsnLysLeuGluAsnLe 605

Qy       1345 AAAAATGATAAGC 1357
Db        605 uLysValIleSer 609

RESULT 23
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 76
;   LENGTH: 708
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Alignment Scores:
Pred. No.:              3.11e-06        Length:         708
Score:                  232.50          Matches:        93
Percent Similarity:     46.27%          Conservative:   118
Best Local Similarity:  20.39%          Mismatches:     161
Query Match:            9.35%           Indels:         84
DB:                     7               Gaps:           19

US-09-837-344-45 (1-1482) x US-11-196-475-76 (1-708)

Qy         49 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA------AAGTTA 102
              :::|||   :::             :::||||||:::::      |||         :::
Db        232 IleGlnGlyGluThrHisLysAlaAspGlnAspLysIleAspThrGluLeuAspAsnIle 251

Qy        103 CAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 162
                 |||      |||::::::   :::            |||   |||::::::|||
Db        252 HisGluSerAspSerAsnIleThrGlu---------ThrIleGluAsnLeuArgAspGln 268

Qy        163 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 222
                          |||             |||             |||   :::|||
Db        269 LeuGluLysAlaThrAspGluGluHisLysLys-----------GluIleGluSerGln 284

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
              ::::::   :::::::::   ||||||:::|||   ::::::      :::|||::::::
Db        285 ValAspAlaLysLysLysGluLysGluGluLeuAspLysLysAlaIleAsnLeuAspLys 304

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA--- 339
                     |||::::::||||||        :::   :::|||:::   :::|||
Db        305 ---------AlaGlnGlnLysLeuAspSerAlaGluAspAsnLeuAspValGlnArgAsp 321

Qy        340 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 399
                  :::||||||:::||||||      :::::::::   :::|||:::
Db        322 ThrValArgGluLysIleGlnGluAspIleAsnGluIleAsnLysGluLysAsnLeuPro 341

Qy        400 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 459
              :::                     ::::::::::::          ||||||   |||
Db        342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy        460 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 519
              ::::::   |||      |||::::::                ::::||   :::|||
Db        362 GluAspLeuGlnGluGlnLeuLysGluAlaGlyAspGluAsnGlnLys---ArgGluIle 380

Qy        520 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              :::      :::|||   ::::::|||   :::|||   |||:::   :::
Db        381 GluLysGlnIleGluIleLysLysArgAspGluGluLeuLeuLysSerLysAspGlyLys 400

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
              :::   ::::::      |||   :::                     |||   |||
Db        401 ValSerLysAspTyrGluAlaLeuAsp-----------------LeuAspArgGluLeu 414

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                 ::::::   :::|||:::      :::   |||:::|||
Db        415 SerLysAlaSerSerLysGluLysSerLysValLysGluGlu----------------- 428

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||          |||:::                  :::|||          |||::::::
Db        429 GluIleThrLysGlyLysSerArgAlaSerLeuGlyAspLeu------AsnAsnAspLys 446

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT------ 813
                    :::|||             ||||||:::|||   :::   :::|||
Db        447 AsnLeuMetLeuProGluAspGlnLysLeuProGluAspLysLysLeuAspSerLysLeu 466

Qy        814 ---------AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATT 864
                      ::::::   |||   :::|||   :::||||||   ::::::   |||
Db        467 AspGlyLysLysGluPheLysProValSerGluValGluLysLeuAspLysIleSerLys 486

Qy        865 ACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
```

```
                :::   |||   |||
Db        487 SerAsnAsnAsnGlu-----------------------------------------Val 492

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAA--------------GATAAATCTGCTGACATA 969
              |||   :::   |||      |||                 |||         ::::::
Db        493 GlyLysLeuSerProLeuAspLysProSerTyrAspAspIleAspSerLysGluGluVal 512

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
                 |||          :::   :::|||:::      ::::::                |||
Db        513 AspAsn---------LysAlaIleAsnLeuGlnLysIleAsp-----------ProLys 525

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                 :::|||:::   ::::::      :::|||      :::|||::::::
Db        526 ValLysAspGlnThrThrSerLeuAsnGluAsp------LeuAspLysAspLeuThrThr 543

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
                    :::|||          |||
Db        544 MetSerIleAspSerSerSerPro---------------------------------- 551

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT------ 1203
                    ::::::   |||:::   |||   |||                 |||
Db        552 ------ValPheLeuGluValIleAspProIleThrAsnLeuGlyThrLeuGlnLeuIle 569

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263
              ||||||:::   |||:::             :::|||   :::::::::   |||::::::
Db        570 AspLeuAsnThrGlyValArg------LeuLysGluSerThrGlnGlnGlyIleGlnArg 587

Qy       1264 GGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG-----------AT 1311
                    ||||||:::   ||||||  :::::::::   :::|||
Db        588 TyrGlyIleTyrGluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLysAl 607

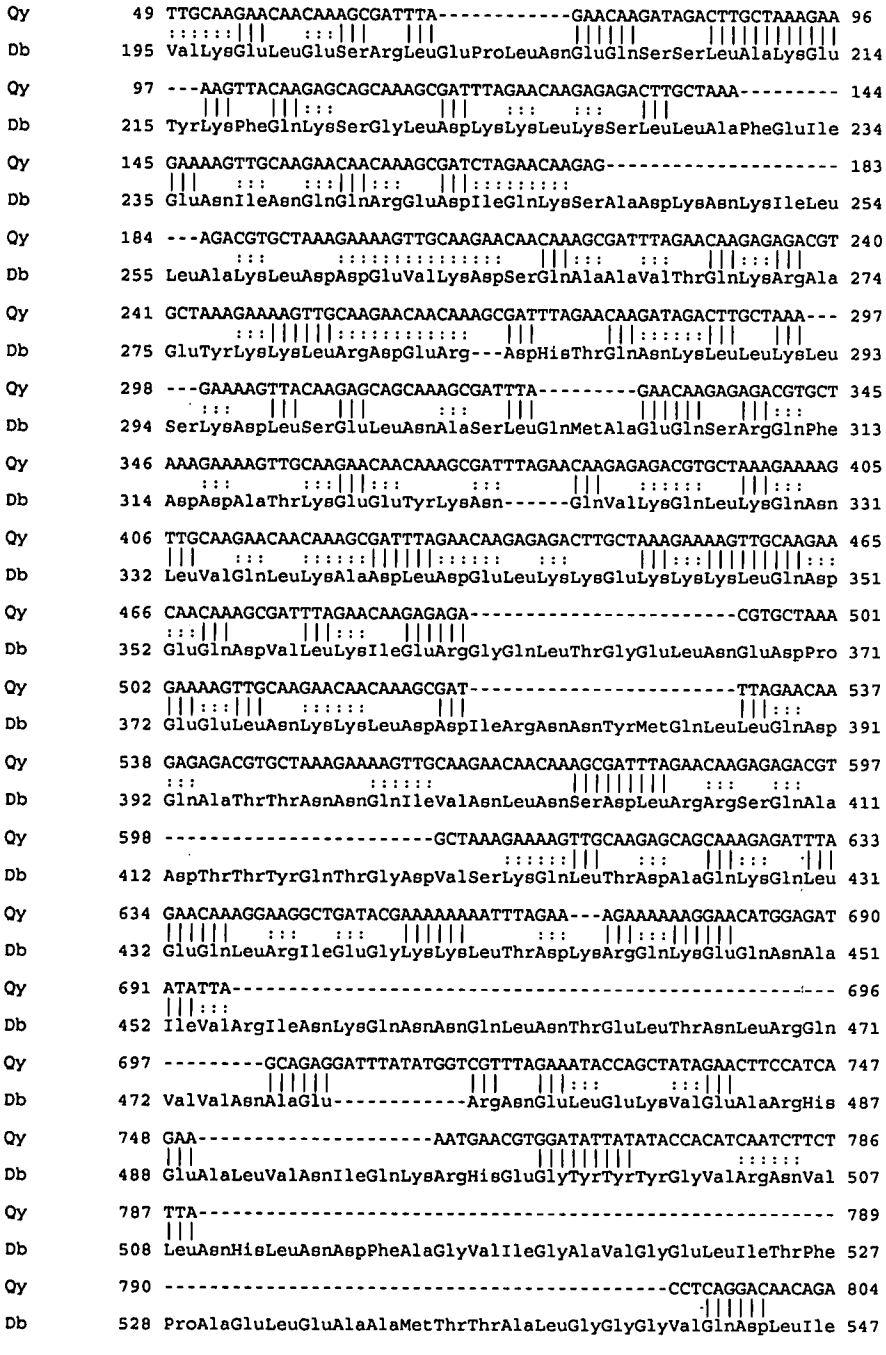
Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
                ||:::   :::::::::   |||:::|||:::|||::::::|||
Db        607 aLysLeuGlnIleLeuAsnLysLeuGluAsnLeuLysValValSer 622

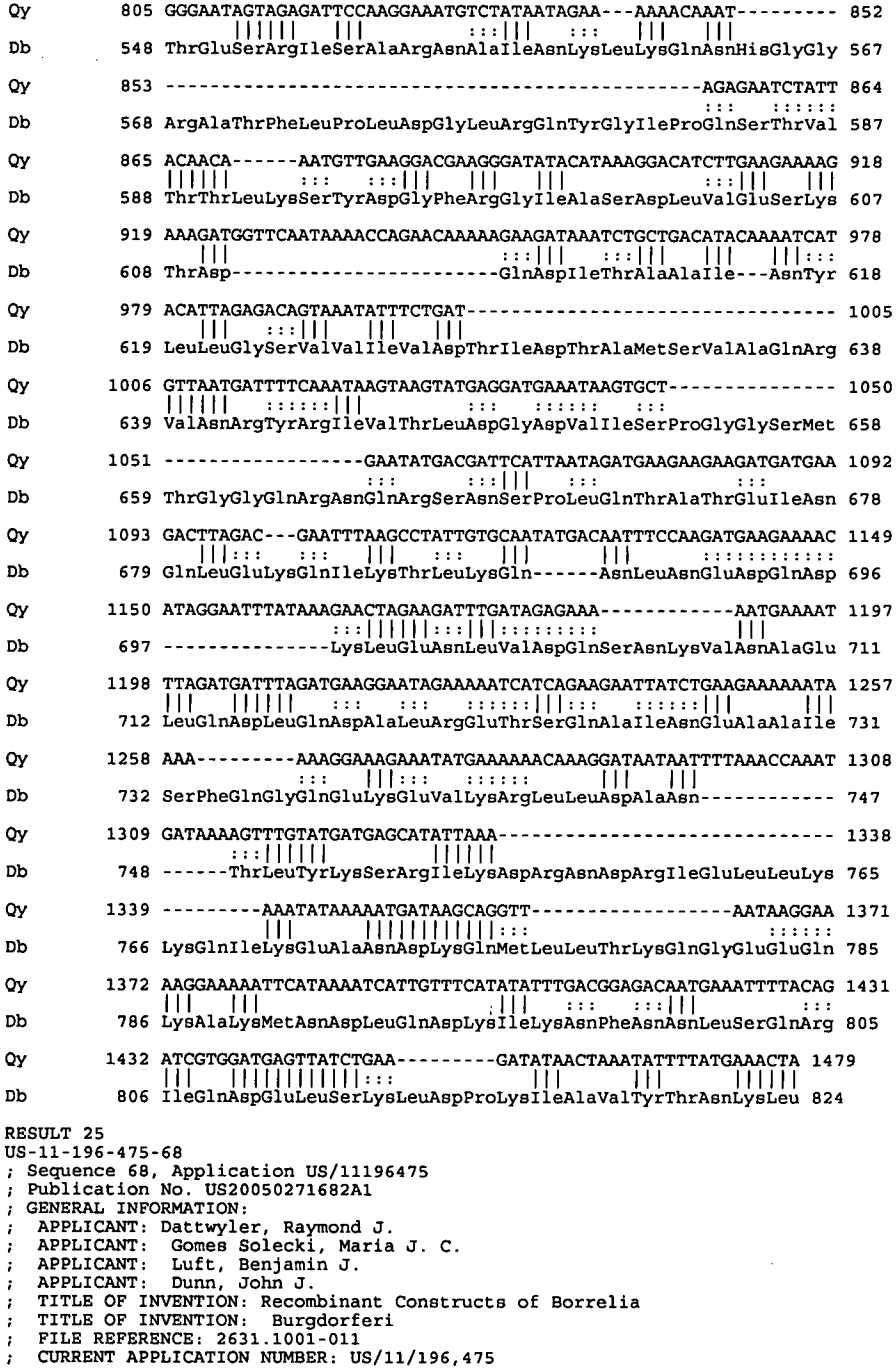
RESULT 24
US-11-074-176-134
; Sequence 134, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
;  APPLICANT: Klaenhammer, Todd R.
;  APPLICANT:  Russell, William M.
;  APPLICANT:  Altermann, Eric
;  APPLICANT:  McAuliffe, Olivia
;  APPLICANT:  Peril, Andrea Azcarate
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding
;  TITLE OF INVENTION:  Stress-Related Proteins and Uses Therefore
;  FILE REFERENCE: 5051-694
;  CURRENT APPLICATION NUMBER: US/11/074,176
;  CURRENT FILING DATE:  2005-03-07
;  PRIOR APPLICATION NUMBER: 60/551,161
;  PRIOR FILING DATE: 2004-03-08
;  NUMBER OF SEQ ID NOS: 381
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 1189
;   TYPE: PRT
;   ORGANISM: Lactobacillus acidophilus
US-11-074-176-134

Alignment Scores:
Pred. No.:              3.23e-06        Length:         1189
Score:                  231.50          Matches:        144
Percent Similarity:     40.21%          Conservative:   129
Best Local Similarity:  21.21%          Mismatches:     191
Query Match:            9.31%           Indels:         215
DB:                     7               Gaps:           33

US-09-837-344-45 (1-1482) x US-11-074-176-134 (1-1189)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGA------------CGTGCTAAAGAAAAG 48
              :::|||   |||:::   |||::::::   :::             |||      :::
Db        175 GluGluAlaGlnAsnGlnLeuLysLysThrGlnAspAsnLeuIleArgIleAsnAspLeu 194
```





RESULT 25
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475

```
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68
;   LENGTH: 693
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-68

Alignment Scores:
Pred. No.:              5.31e-06        Length:         693
Score:                  228.00          Matches:        98
Percent Similarity:     43.82%          Conservative:   111
Best Local Similarity:  20.55%          Mismatches:     143
Query Match:            9.17%           Indels:         125
DB:                     7               Gaps:           19

US-09-837-344-45 (1-1482) x US-11-196-475-68 (1-693)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
Db        230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy         73 GAACAAGATAGACTTGCTAAA----------------GAAAAGTTACAAGAGCAGCAA 114
Db        247 GluLeuAspAsnValHisLysSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        267 AspGlnLeuGlu---LysAlaThrAspGluGluHisArgLysGluIleGluSerGlnVal 285

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
Db        286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGA---CTT 291
Db        304 ---LysAlaGlnGlnLysLeuAspSerSerGluAspAsnLeuAspIleGlnArgAspThr 322

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
Db        323 ValArgGluLysIleGlnGluAspIleAspGluIleAsnLysGluLysAsnLeuProLys 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
Db        343 ProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeuGlu 362

Qy        412 GAACAACAA----------------AGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 453
Db        363 AspLeuGlnGluGlnLeuLysGluThrSerAspGluAsnGlnLysArgGluIleGluLys 382

Qy        454 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 513
Db        383 GlnIleGluIleLysLysSerAspGluGluLeuLeuLysSerLysAspProLysAlaLeu 402

Qy        514 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 573
Db        403 AspLeuAsnGlyAspLeuAsnSerLysValSerSerLysGluLysIleLysGlyLysGlu 422

Qy        574 AGCGATTTA---GAACAAGAGAGACGTGCTAAA----------------------GAAAAG 609
Db        423 GlyGluIleValLysGluGluSerLysAlaSerLeuAlaAspLeuAsnAsnAspGluAsn 442

Qy        610 TTGCAA------GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAAT 663
Db        443 LeuMetArgProGluAspGlnLysLeuSerGluAspLysLysLeuAspSerLysLysAsn 462

Qy        664 TTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAA 723
Db        463 LeuLys---------------------------------------------------- 464

Qy        724 ATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCT 783
Db        465 ---ProValSerGluIleGluArgValAsnGlu-------------------------- 474

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAA 843
Db        475 ------------IleSerLysSerAsnAsnAsnGluIleSerGluSerSerProLeuTyr 490

Qy        844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903
Db        491 LysProSer------------TyrSerAspMetAspSerLysGluGlyIle--------- 503

Qy        904 CATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
Db        504 ---------AspAsnLysAspValAsnLeuGlnGluThrLysSerGlnThrLysSerGln 520

Qy        964 GACATACAA---AATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db        521 ProThrSerLeuAsnGlnAspLeuThrThrMetSerIleAspSerSerAsnProVal--- 539

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db        539 ---------------------------------------------------------- 539

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db        540 ------------PheLeuGluValIleAspProIleThr---------------------- 548

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db        549 ------AsnLeuGlyThrLeuGln--------------------------------Leu 555

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
Db        556 IleAspLeuAsnThrGlyValArg------LeuLysGluSerThrGlnGlnGlyIleGln 573

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG---------- 1309
Db        574 ArgTyrGlyIleTyrGluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLy 593

Qy       1310 -ATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
Db        593 sAlaLysLeuGlnIleLeuAsnLysLeuGluAsnLeuLysValIleSer 609

RESULT 26
US-10-821-234-901
; Sequence 901, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 901
;   LENGTH: 1586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
```

US-10-821-234-901

Alignment Scores:

Pred. No.:	4.67e-06	Length:	1586
Score:	228.00	Matches:	126
Percent Similarity:	38.30%	Conservative:	95
Best Local Similarity:	21.84%	Mismatches:	188
Query Match:	9.17%	Indels:	168
DB:	6	Gaps:	23

US-09-837-344-45 (1-1482) x US-10-821-234-901 (1-1586)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAA 54
              ::::::|||:::      ||||||   :::      |||   ||||||      ||||||
Db       1025 GluGlnGlnArgLysAlaLeuGluAlaLysAlaAlaAlaPheGluLysGlnValLeuGln 1044

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG------ 108
                 |||   |||       |||   :::          :::::::::|||   |||
Db       1045 LeuGlnAlaSerHisArgGluSerGluGluAlaLeuGlnLysArgLeuAspGluValSer 1064

Qy        109 ---------------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 153
                           ||||||             |||         ||||||
Db       1065 ArgGluLeuCysHisThrGlnSerSerHisAlaSerLeuArgAlaAspAlaGluLysAla 1084

Qy        154 CAAGAACAACAAAGCGATCTA----------------------------GAACAAGAG 183
              ||||||||||||      :::                          |||   |||
Db       1085 GlnGluGlnGlnGlnGlnMetAlaGluLeuHisSerLysLeuGlnSerSerGluAlaGlu 1104

Qy        184 AGACGTGCTAAA---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
                 |||:::|||   |||:::|||                |||::::::   |||
Db       1105 ValArgSerLysCysGluGluLeuSerGlyLeuHisGlyGlnLeuGlnGluAlaArgAla 1124

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA-----------------CAA 282
                    :::|||   |||:::         :::|||                    |||
Db       1125 GluAsnSerGlnLeuThrGluArgIleArgSerIleGluAlaLeuLeuGluAlaGlyGln 1144

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
                 |||   |||::::::   :::|||       |||::::::   :::|||::::::
Db       1145 AlaArgAspAlaGlnAsp---ValGlnAlaSerGlnAlaGluAlaAspGlnGlnGlnThr 1163

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-----------AGA 390
                   ::::::|||:::   |||   |||   ||||||:::|||              |||
Db       1164 ArgLeuLysGluLeuGluSerGlnValSerGlyLeuGluLysGluAlaIleGluLeuArg 1183

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA--------------- 435
                 |||   |||:::   :::   :::   :::||||||   :::
Db       1184 GluAlaValGluGlnGlnLysValLysAsnAsnAspLeuArgGluLysAsnTrpLysAla 1203

Qy        436 --------------------GAGAGACTTGCTAAAGAAAAGTTGCAAGAA--------- 465
                                 |||:::      ||||||||||||
Db       1204 MetGluAlaLeuAlaThrAlaGluGlnAlaCysLysGluLysLeuHisSerLeuThrGln 1223

Qy        465 ------------------------------------------------------------ 465

Db       1224 AlaLysGluGluSerGluLysGlnLeuCysLeuIleGluAlaGlnThrMetGluAlaLeu 1243

Qy        466 -------------------------------CAACAAAGCGATTTAGAACAAGAGAGA 492
                                              ||||||:::      |||      :::
Db       1244 LeuAlaLeuLeuProGluLeuSerValLeuAlaGlnGlnAsnTyrThrGluTrpLeuGln 1263

Qy        493 CGTGCTAAAGAAAAG-----------------------------TTGCAAGAACAACAA 522
                 |||||||||                               |||
Db       1264 AspLeuLysGluLysGlyProThrLeuLeuLysHisProProAlaProAlaGluProSer 1283

Qy        523 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---CAAAGCGAT 579
                 |||||||||      :::   |||      |||:::   |||      |||::::::
Db       1284 SerAspLeuAlaSerLysLeuArgGluAlaGluGluThrGlnSerThrLeuGlnAlaGlu 1303

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                 :::|||   |||            :::   :::          |||||||||   |||
Db       1304 CysAspGlnTyrArgSerIleLeuAlaGluThrGluGlyMetLeuArgAspLeu---Gln 1322

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAG---------------------- 678
```

```
                 :::      :::   ::::::   :::   |||   |||
Db       1323 LysSerValGluGluGluGluGlnValTrpArgAlaLysValGlyAlaAlaGluGluGlu 1342

Qy        679 ------------------------GAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                                      :::|||   :::   :::   |||   |||   |||
Db       1343 LeuGlnLysSerArgValThrValLysHisLeuGluGluIleValGluLysLeuLysGly 1362

Qy        715 CGTTTAGAAATACCA-----------------GCTATAGAACTTCCATCAGAAAATGAA 756
                 ||||||                        :::   |||  :::|||   |||
Db       1363 GluLeuGluSerSerAspGlnValArgGluHisThrLeuHisLeuGluAlaGluLeuGlu 1382

Qy        756 ------------------------------------------------------------ 756

Db       1383 LysHisMetAlaAlaAlaSerAlaGluCysGlnAsnTyrAlaLysGluValAlaGlyLeu 1402

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
                 |||          :::      ||||||   |||                    :::
Db       1403 ArgGlnLeuLeuLeuGluSerGlnSerGlnLeuAspAlaAlaLysSerGluAlaGlnLys 1422

Qy        817 GATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTT 876
                 |||   |||::::::::::::   :::          |||
Db       1423 GlnSerAspGluLeuAlaLeuValArgGlnGlnLeuSerGluMet--------------- 1437

Qy        877 GAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA--- 933
                                 |||   |||:::|||          ||||||   |||
Db       1438 ---------------------LysSerHisValGlu---------AspGlyAspIleAla 1447

Qy        934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                              ||||||                ::::::|||
Db       1448 GlyAlaProAlaSerSerProGluAlaProProAlaGluGlnAspProValGlnLeuLys 1467

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
                 |||                :::   :::   |||   |||      :::
Db       1468 ThrGlnLeuGluTrpThrGluAlaIleLeuGluAspGluGlnThrGlnArg------Gln 1485

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTA 1098
                 :::::::::||||||::::::::::::                    |||:::|||
Db       1486 LysLeuThrAlaGluPheGluGluAlaGlnThrSerAlaCysArgLeuGlnGluGluLeu 1505

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
                 ::::::   :::                ::::::   :::   ||||||
Db       1506 GluLysLeuArgThrAlaGlyProLeuGluSerSerGluThrGluGlu------------ 1521

Qy       1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------ 1206
                    :::|||::::::   :::||||||   :::   |||
Db       1522 AlaSerGlnLeuLysGluArgLeuGluLysGluLysLysLeuThrSerAspLeuGlyArg 1541

Qy       1207 ------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
                         |||   |||   ::::::   :::   |||:::|||:::   ||||||
Db       1542 AlaAlaThrArgLeuGlnGluLeuLeuLysThrThrGlnGluGlnLeuAlaArgGluLys 1561

Qy       1255 ------ATAAAAAAA---GGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
                    :::||||||      ::::::   ||||||   :::|||   :::
Db       1562 AspThrValLysLysLeuGlnGluGlnLeuGluLysAlaGluAspGlySer 1578
```

RESULT 27
US-10-878-556A-136
; Sequence 136, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression
; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28
; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 136
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: sw_hum/rrb1_human

```
;   DATABASE ENTRY DATE: 2003-02-28
US-10-878-556A-136

Alignment Scores:
Pred. No.:                8.57e-06        Length:          1410
Score:                    223.00          Matches:         122
Percent Similarity:       39.13%          Conservative:    103
Best Local Similarity:    21.22%          Mismatches:      186
Query Match:              8.97%           Indels:          164
DB:                       6               Gaps:            23

US-09-837-344-45 (1-1482) x US-10-878-556A-136 (1-1410)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG------TTGCAA 54
              ::::::|||:::     ||||||   :::      |||   ||||||      ||||||
Db        849 GluGlnGlnArgLysAlaLeuGluAlaLysAlaAlaAlaPheGluLysGlnValLeuGln 868

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG------ 108
               |||   |||      |||   :::         :::::::::|||   |||
Db        869 LeuGlnAlaSerHisArgGluSerGluGluAlaLeuGlnLysArgLeuAspGluValSer 888

Qy        109 ---------------CAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG 153
                           ||||||                 |||          ||||||
Db        889 ArgGluLeuCysHisThrGlnSerSerHisAlaSerLeuArgAlaAspAlaGluLysAla 908

Qy        154 CAAGAACAACAAAGCGATCTA----------------------------GAACAAGAG 183
              ||||||||||||     :::                            |||   |||
Db        909 GlnGluGlnGlnGlnGlnMetAlaGluLeuHisSerLysLeuGlnSerSerGluAlaGlu 928

Qy        184 AGACGTGCTAAA-----------------------GAAAAGTTGCAAGAACAACAAAGC 219
              |||:::|||                        :::||||||||||   ::::::
Db        929 ValArgSerLysCysGluGluLeuSerGlyLeuHisGlyGlnLeuGlnGluAlaArgAla 948

Qy        220 GATTTAGAACAA-----------------------------------------------GAG 234
              :::      |||                                               :::
Db        949 GluAsnSerGlnLeuThrGluArgIleArgSerIleGluAlaLeuLeuGluAlaGlyGln 968

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 294
              |||      :::   :::|||      |||::::::   :::|||   :::
Db        969 AlaArgAspAlaGlnAspValGlnAlaSerGlnAlaGluAlaAspGlnGlnGlnThrArg 988

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG------------AGACGT 342
              ::::::|||:::   |||   |||   ||||||:::|||              |||
Db        989 LeuLysGluLeuGluSerGlnValSerGlyLeuGluLysGluAlaIleGluLeuArgGlu 1008

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA----------------- 384
              |||   |||:::   :::   :::   :::||||||   :::
Db       1009 AlaValGluGlnGlnLysValLysAsnAsnAspLeuArgGluLysAsnTrpLysAlaMet 1028

Qy        385 ------------------GAGAGACGTGCTAAAGAAAAGTTG------------------ 408
                              |||:::   ||||||||||||
Db       1029 GluAlaLeuAlaThrAlaGluGlnAlaCysLysGluLysLeuLeuSerLeuThrGlnAla 1048

Qy        409 ---CAAGAACAACAAAGCGATTTA---------------------------------- 429
                 :::|||   :::      |||
Db       1049 LysGluGluSerGluLysGlnLeuCysLeuIleGluAlaGlnThrMetGluAlaLeuLeu 1068

Qy        430 ------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
                        |||          ||||||::::::          |||          |||
Db       1069 AlaLeuLeuProGluLeuSerValLeuAlaGlnGlnAsnTyrThrGluTrpLeuGlnAsp 1088

Qy        478 TTAGAACAAGAG---------------------------------------------- 489
              |||::::::::::
Db       1089 LeuLysGluLysGlyProThrLeuLeuLysHisProProAlaProAlaGluProSerSer 1108

Qy        490 -----------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531
                              |||   |||:::|||               |||::::::
Db       1109 AspLeuAlaSerLysLeuArgGluAlaGluGluThrGlnSerThrLeuGlnAlaGluCys 1128

Qy        532 GAACAAGAGAGA------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
              :::|||   |||                  :::      |||::::::   |||
Db       1129 AspGlnTyrArgSerIleLeuAlaGluThrGluGlyMetLeuArgAspLeuGlnLysSer 1148

Qy        580 TTAGAACAAGAGAGACGT----------------------GCTAAAGAAAAGTTGCAA 615
              :::|||:::|||   :::                     |||:::|||:::||||||
Db       1149 ValGluGluGluGluGlnValTrpArgAlaLysValGlyAlaAlaGluGluGluLeuGln 1168

Qy        616 GAGCAGCAA---------AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTA 666
              :::   :::         :::   ||||||:::                  |||      |||
Db       1169 LysSerArgValThrValLysHisLeuGluGluIleValGluLysLeuLysGlyGluLeu 1188

Qy        667 GAA---------AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT 717
              |||         :::   :::||||||         |||      :::|||
Db       1189 GluSerSerAspGlnValArgGluHisThrSerHisLeuGluAlaGluLeuGluLysHis 1208

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA---------------CGTGGA 762
              :::         :::   |||      :::      :::                 |||
Db       1209 MetAlaAlaAlaSerAlaGluCysGlnAsnTyrAlaLysGluValAlaGlyLeuArgGln 1228

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC 822
              :::      ||||||   |||                              :::   |||
Db       1229 LeuLeuLeuGluSerGlnSerGlnLeuAspAlaAlaLysSerGluAlaGlnLysGlnSer 1248

Qy        823 AAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGA 882
              |||:::::::::::::   :::      |||
Db       1249 AspGluLeuAlaLeuValArgGlnGlnLeuSerGluMet--------------------- 1261

Qy        883 CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA--------- 933
                          |||   |||:::|||         ||||||   |||
Db       1262 --------------LysSerHisValGlu---------AspGlyAspIleAlaGlyAla 1273

Qy        934 ---------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
                        ||||||            :::         :::|||      |||
Db       1274 ProAlaSerSerProGluAlaProProAlaGluGlnAspProValGlnLeuLysThrGln 1293

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                          :::   :::   |||   |||      :::      ::::::
Db       1294 LeuGluTrpThrGluAlaIleLeuGluAspGluGlnThrGlnArg------GlnLysLeu 1311

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
              :::||||||:::::::::::                        |||:::|||::::::
Db       1312 ThrAlaGluPheGluGluAlaGlnThrSerAlaCysArgLeuGlnGluGluLeuGluLys 1331

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
                 :::                 ::::::   :::   ||||||
Db       1332 LeuArgThrAlaGlyProLeuGluSerSerGluThrGluGlu------------AlaSer 1347

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------------ 1206
              :::|||::::::   :::||||||   :::   |||
Db       1348 GlnLeuLysGluArgLeuGluLysGluLysLysLeuThrSerAspLeuGlyArgAlaAla 1367

Qy       1207 ------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA------ 1254
                    |||   |||   ::::::   :::   |||:::|||:::   ||||||
Db       1368 ThrArgLeuGlnLeuLeuLysThrThrGlnGluGlnLeuAlaArgGluLysAspThr 1387

Qy       1255 ATAAAAAAA---GGAAAGAAATATGAAAAAACAAAGGATAATAAT 1296
              :::||||||      ::::::   ||||||   :::|||   :::
Db       1388 ValLysLysLeuGlnGluGlnLeuGluLysAlaGluAspGlySer 1402

RESULT 28
US-10-821-234-975
; Sequence 975, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
```

```
; SEQ ID NO 975
;   LENGTH: 989
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-975

Alignment Scores:
Pred. No.:                1.22e-05        Length:          989
Score:                    220.50          Matches:         116
Percent Similarity:       41.77%          Conservative:    115
Best Local Similarity:    20.98%          Mismatches:      210
Query Match:              8.87%           Indels:          112
DB:                       6               Gaps:            21

US-09-837-344-45 (1-1482) x US-10-821-234-975 (1-989)

Qy         31 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 90
              ::::::   :::|||   ||||||:::   |||      ||||||
Db        441 GlnGlnLeuGlnGluIleLeuGlnAspLeuGlnLysArgLeuGluSerSerGluAlaGlu 460

Qy         91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA------CTTGCTAAA 144
              :::::::::||||||   :::                |||          |||
Db        461 ArgLysGlnLeuGlnValGluLeuGlnSerArgArgAlaGluLeuValCysLeuAsnAsn 480

Qy        145 GAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 204
              ::::::   |||      |||||||||   |||                    ||||||
Db        481 ThrGluIleSerGluAsnSerSerAspLeuSerGln----------------LysLeu 494

Qy        205 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 264
              :::|||   ||||||      |||:::      :::                :::|||:::|||
Db        495 LysGluThrGlnSerLysTyrGluGluAlaMetLysGluValLeuSerValGlnLysGln 514

Qy        265 CAA-----------------------------------AGCGATTTAGAACAAGAT 285
                                                    |||          :::
Db        515 MetLysLeuGlyLeuValSerProGluSerMetAspAsnTyrSerHisPheHisGluLeu 534

Qy        286 AGACTTGCTAAAGAAAAGTTA------------CAAGAGCAGCAAAGCGATTTAGAACAA 333
              |||:::   :::|||::::::           |||:::   |||:::   ||||||:::
Db        535 ArgValThrGluGluGluIleAsnValLeuLysGlnAspLeuGlnAsnAlaLeuGluGlu 554

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 393
                 |||   |||||||||:::::|||   :::      |||   :::      :::
Db        555 SerGluArgAsnLysGluLysValArgGluLeuGluGluLysLeuValGluArgGluLys 574

Qy        394 GCTAAA-----------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
                                    |||:::   :::|||   :::|||
Db        575 GlyThrValIleLysProProValGluGluTyrGluGluMetLysSerSerTyrCysSer 594

Qy        436 ------GAGAGACTTGCTAAAGAAAAG---------TTGCAAGAACAACAAAGCGATTTA 480
                    |||   :::   |||||||||           ::::::   |||
Db        595 ValIleGluAsnMetAsnLysGluLysAlaPheLeuPheGluLysTyrGlnGluAlaGln 614

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||:::      :::   |||:::   |||:::   |||   :::
Db        615 GluGluIleMetLysLeuLysAspThrLeuLysSerGlnMetThr-------------- 629

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
                  :::|||          |||   :::          |||          |||
Db        630 ---------GlnGluAlaSerAspGluAlaGluAspMetLysGluAlaMetAsnArgMet 646

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
                 ::::::|||   :::|||      :::|||   |||      :::   :::
Db        647 IleAspGluLeuAsnLysGlnValSerGluLeuSerGlnLeuTyrLysGluAlaGlnAla 666

Qy        661 AATTTAGAA---------AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATAT 711
              ||||||          :::::::|||          |||:::   ||||||   ::::::
Db        667 GluLeuGluAspTyrArgLysArgLysSerLeuGluAspValThrAlaGluTyrIleHis 686

Qy        712 GGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATA 771
                                          :::|||:::|||:::         :::
Db        687 ----------------------------LysAlaGluHisGluLysLeuMetGlnLeu 696

Qy        772 CCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC--------- 822
```

```
                 :::   |||          ::::::      :::         |||
Db        697 ThrAsnValSerArgAlaLysAlaGluAspAlaLeuSerGluMetLysSerGlnTyrSer 716

Qy        823 --------------------------AAGGAAATGTCTATAATAGAAAAAACAAATAGA 855
                                         |||::::::        :::|||   |||
Db        717 LysValLeuAsnGluLeuThrGlnLeuLysGlnLeuValAspAlaGlnLysGluAsnSer 736

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA--- 912
                 |||||||||   :::::::::          :::                 |||
Db        737 ValSerIleThrGluHisGluGlnValIleThrThrLeuArgThrAlaAlaLysGluMet 756

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAA------CAAAAAGAAGATAAATCTGCTGAC 966
              |||:::|||      ::::::|||                |||||   :::   |||
Db        757 GluGluLysIleSerAsnLeuLysGluHisLeuAlaSerLysGluValGluValAlaLys 776

Qy        967 ATACAAAATCATACATTAGAG----------------------ACAGTAAATATTTCT 1002
              ::::::          |||||                     |||      |||
Db        777 LeuGluLysGlnLeuLeuGluGluLysAlaAlaMetThrAspAlaMetValProArgSer 796

Qy       1003 GATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT---------GCTGAA 1053
                        |||   |||      |||   |||:::|||          ::::::
Db        797 SerTyrGluLysLeuGln---SerSerLeuGluSerGluValSerValLeuAlaSerLys 815

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
                   :::|||:::   ::::::|||:::          ::::::   :::   :::
Db        816 LeuLysGluSerValLysGluLysGluLysValHisSerGluValValGlnIleArgSer 835

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173
                |||                :::::::::||||||||
Db        836 GluValSerGlnValLys---ArgGluLysGluAsnIleGlnThrLeu------------ 850

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233
                |||      |||   :::   :::::::::|||   :::      :::::::::
Db        851 ---LeuLysSerLysGluGlnGluValAsnGluLeuLeuGlnLysPheGlnGlnAlaGln 869

Qy       1234 GAAGAATTATCTGAAGAAAAA------ATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
              |||||||||:::|||   |||         :::         |||   |||:::   |||
Db        870 GluGluLeuAlaGluMetLysArgTyrAlaGluSerSerSerLysLeuGluGluAspLys 889

Qy       1288 GATAATAATTTTAAACCAAATG----------------ATAAAAGTTTGTATGATGAG 1329
              |||        :::|||:::|||                :::|||      :::   ||
Db        890 AspLysLys-IleAsnGluMetSerLysGluValThrLysLeuLysGluAlaLeuAsnSe 909

Qy       1330 CATA-----------------------------------TTAAAAAATATAAAAAT 1350
              |:::                                  :::   ::::::
Db        909 rLeuSerGlnLeuSerTyrSerThrSerSerSerLysArgGlnSerGlnGlnLeuGluAl 929

Qy       1351 GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGAC 1410
                :::   :::   :::::::::   :::|||            |||
Db        929 aLeuGlnGlnGlnValLysGlnLeuGlnAsnGlnLeuAlaGluCysLysLysGlnHisGl 949

Qy       1411 GGAGACAATGAAATTTTACAGATCGTGGATGAGTTAT 1447
                |||   :::      ||||||      :::   |||
Db        949 nGluValIleSerValTyrArgMetHisLeuLeuTyr 961

RESULT 29
US-11-196-475-74
; Sequence 74, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
;  APPLICANT: Dattwyler, Raymond J.
;  APPLICANT:  Gomes Solecki, Maria J. C.
;  APPLICANT:  Luft, Benjamin J.
;  APPLICANT:  Dunn, John J.
;  TITLE OF INVENTION: Recombinant Constructs of Borrelia
;  TITLE OF INVENTION:  Burgdorferi
;  FILE REFERENCE: 2631.1001-011
;  CURRENT APPLICATION NUMBER: US/11/196,475
;  CURRENT FILING DATE:  2005-08-03
;  PRIOR APPLICATION NUMBER: US 08/148,191
;  PRIOR FILING DATE: 1993-11-01
;  PRIOR APPLICATION NUMBER: US 08/235,836
;  PRIOR FILING DATE: 1994-04-29
```

```
;  PRIOR APPLICATION NUMBER: US 09/666,017
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: US 60/226,484
;  PRIOR FILING DATE: 2000-08-18
;  PRIOR APPLICATION NUMBER: PCT/US01/24736
;  PRIOR FILING DATE: 2001-08-07
;  NUMBER OF SEQ ID NOS: 213
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 700
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Alignment Scores:
Pred. No.:                1.83e-05       Length:         700
Score:                    217.50         Matches:        102
Percent Similarity:       46.19%         Conservative:   116
Best Local Similarity:    21.61%         Mismatches:     144
Query Match:              8.75%          Indels:         110
DB:                       7              Gaps:           25

US-09-837-344-45 (1-1482) x US-11-196-475-74 (1-700)

Qy         13 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 72
Db        230 ThrAspIleGlnGlyGlu---------ThrHisLysAlaAspGlnAspLysIleAspIle 246

Qy         73 GAACAAGATAGACTTGCTAAA----------------GAAAAGTTACAAGAGCAGCAA 114
Db        247 GluLeuAspAsnIleHisGluSerAspSerAsnIleThrGluThrIleGluAsnLeuArg 266

Qy        115 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTA 174
Db        267 AspGlnLeuGlu---LysAlaThrAspGluGluHisLysLysGluIleGluSerGlnVal 285

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
Db        286 AspAlaLysLysLysGlnLysGluGluLeuAspLysLysAlaIleAspLeuAsp------ 303

Qy        235 AGACGTGCTAAAGAAAAGTTGCAA------GAACAACAAAGCGATTTAGAACAAGATAGA 288
Db        304 ---LysAlaGlnGlnLysLeuAspPheAlaGluAspAsnLeuAspIleGlnArgAspThr 322

Qy        289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
Db        323 Val---ArgGluLysLeuGlnGluAsnIleAsnGluThrAsnLysGluLysAsnLeuPro 341

Qy        349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
Db        342 LysProGlyAspValSerSerProLysValAspLysGlnLeuGlnIleLysGluSerLeu 361

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
Db        362 GluAspLeuGlnGluGlnLeu--------------------------------LysGlu 370

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
Db        371 ThrGlyAspGluAsnGlnLysArgGluIleGluLysGlnIleGluIleLysLysSerAsp 390

Qy        529 ------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
Db        391 GluLysLeuLeuLysSerLysAspAspLysAlaSerLysAspGlyLysAlaLeuAspLeu 410

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
Db        411 AspArgGluLeuAsnSerLysAlaSerSerLysGluLysSerLysAlaLysGluGluGlu 430

Qy        643 ------AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696
Db        431 IleThrLysGlyLysSerGlnLysSerLeu-----------------GlyAspLeu--- 443

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756
Db        444 ----------------------------------------------------AsnAsnAsp 446

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT--- 813
Db        447 GluAsnLeuMetMetProGluAspGlnLysLeuProGluValLysLysLeuAspSerLys 466

Qy        814 AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
Db        467 LysGluPheLysProValSerGluValGluLysLeuAspLys----------------- 480

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db        481 ------------------IlePheLysSerAsn---------AsnAsnValGlyGluLeu 491

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCT------GACATACAAAATCATACATTAGAG 987
Db        492 SerPro---------LeuAspLysSerSerTyrLysAspIleAspSerLysGlu---Glu 507

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAGGATGAA 1041
Db        508 ThrValAsn---LysAspValAsn---LeuGlnLysThrLysProGlnValLysAspGln 525

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
Db        526 ValThrSer-----------------LeuAsnGluAspLeuThrThrMetSerIleAsp 539

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        540 SerSerSerPro-----------------------------------------ValPhe 545

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT------GATTTAGATGAA 1215
Db        546 LeuGluValIleAspProIleThrAsnLeuGlyThrLeuGlnLeuIleAspLeuAsnThr 565

Qy       1216 GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
Db        566 GlyValArg------LeuLysGluSerThrGlnGlnGlyIleGlnArgTyrGlyIleTyr 583

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATG------------ATAAAAGTTTGTAT 1323
Db        584 GluArgGluLysAsp-LeuValValIleLysMetAspSerGlyLysAlaLysLeuGlnIl 603

Qy       1324 GATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
Db        603 eLeuAspLysLeuGluAsnLeuLysValValSer 614

RESULT 30
US-10-821-234-1050
; Sequence 1050, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1050
;   LENGTH: 1410
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1050

Alignment Scores:
Pred. No.:                2.08e-05       Length:         1410
Score:                    215.50         Matches:        85
```

Percent Similarity: 48.82% Conservative: 60
Best Local Similarity: 28.62% Mismatches: 107
Query Match: 8.67% Indels: 45
DB: 6 Gaps: 14

US-09-837-344-45 (1-1482) x US-10-821-234-1050 (1-1410)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::::::::::   ||||||:::||||||      |||   :::      ||||||
Db       1086 GluLysGluLysAlaArgLeuGluGluGluArgGluArgLysArgArg------GluGln 1103

Qy         61 CAAAGCGATTTAGAACAAGATAGA---------------------------------- 84
              :::      |||   |||   |||
Db       1104 GluArgLysLeuGlyGlnMetArgThrGlnSerLeuAsnProAlaProPheSerProLeu 1123

Qy         85 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA------CAAGAGAGACTT 138
                 |||::::::         |||:::   |||   |||:::      ||||||   :::
Db       1124 ThrAlaGlnGlnMetLysProGluLysProSerThrLeuGlnArgProGlnGluThrVal 1143

Qy        139 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA 198
                ::::||   ||||||   ||||||                |||||||||      :::
Db       1144 IleArgGlu---LeuGlnProGlnGlnGlnProArgThrIleGluArgArgAspLeuGln 1162

Qy        199 AAGTTGCAAGAACAACAAAGCGATTTAGAACAA-------------------------- 231
                :::         :::   :::|||
Db       1163 TyrIleThrValSerLysGluGluLeuSerSerGlyThrValCysProProAspProTrp 1182

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
              :::|||   |||||||||||||||:::::||||||                |||   |||
Db       1183 LysArgAspAlaLysGluLysLeuGluLysGlnGlnGlnMetHisIleValAspMetLeu 1202

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
              :::||||||   :::||||||   ||||||      ::::::                :::
Db       1203 SerLysGlu---IleGlnGluLeuGlnSerLysProAspArgSerAlaGluGluSerAsp 1221

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 411
              :::|||::::::         :::   :::   ::::::|||   :::|||      |||
Db       1222 ArgLeuArgLysLeuMetLeuGluTrpGlnPheGlnLysArgLeuGlnGluSerLysGln 1241

Qy        412 GAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 471
              :::   :::   |||   |||:::|||           :::   :::         |||
Db       1242 LysAspGluAspAspGluGluGluGluAspAspAspValAspThrMetLeuIleMetGln 1261

Qy        472 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA---CAACAAAGCGAT 528
                   ||||||   ||||||||||||:::      :::                |||
Db       1262 Arg---LeuGluAlaGluArgArgAlaArgThrAlaMetProAlaIleSerValLeuAsp 1280

Qy        529 TTA------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
              |||      ::::::|||||||||   :::::::::|||:::|||   :::
Db       1281 LeuLeuGlnAspGluGluArgArgArgGlnGlnGlnLeuGluGluMetArgLys---Arg 1299

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
              |||   |||   ||||||:::::::::   :::|||:::::::|||       |||:::|||
Db       1300 GluAlaGluAspArgAlaArgGlnGlu---GluGluArgArgArgGlnGluGluGluArg 1318

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
                    ||||||::::::   |||   |||:::                      |||
Db       1319 ---------ThrLysArgAspAlaGluGluLysArgArgGln------------GluGlu 1331

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
                 |||   |||||||||                                |||      |||
Db       1332 GlyTyrTyrSerArgLeuGluAlaGluArgArgArgGlnHisAspGluAlaAlaArgArg 1351

Qy        763 TATTATATACCACAT--------------CAATCTTCTTTACCTCAGGAC 798
                    |||                  :::      ||||||:::|||
Db       1352 LeuLeuGluProGluAlaProGlyLeuCysArgProProLeuProArgAsp 1368
```

Search completed: December 21, 2005, 11:03:38
Job time : 111.432 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:27:59 ; Search time 8587.96 Seconds
(without alignments)
9809.312 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1482	100.0	1482	6	AR162974	AR162974 Sequence
2	1482	100.0	1482	6	AR178263	AR178263 Sequence
3	1480.4	99.9	1482	6	AR162970	AR162970 Sequence
4	1480.4	99.9	1482	6	AR178259	AR178259 Sequence
5	1480.4	99.9	1496	2	PFLSA13	Z30320 P.falciparu
6	1444.4	97.5	1493	6	A28743	A28743 cDNA for LS
7	1444.4	97.5	1493	6	AR162967	AR162967 Sequence
8	1444.4	97.5	1493	6	AR178256	AR178256 Sequence
9	1440.4	97.2	5970	2	PFLSA1G	X56203 P.falciparu
10	1405.2	94.8	253001	2	AE014834	AE014834 Plasmodiu
11	948.4	64.0	950	6	A28740	A28740 cDNA for LS
12	948.4	64.0	950	6	AR162961	AR162961 Sequence
13	948.4	64.0	950	6	AR178250	AR178250 Sequence
14	841.4	56.8	843	2	PFALSAC	L40885 Plasmodium
15	839.8	56.7	843	2	PFALSAB	L40884 Plasmodium
16	839.8	56.7	843	2	PFALSAG	L40889 Plasmodium
17	839.8	56.7	843	2	PFALSAH	L40890 Plasmodium
18	839.8	56.7	843	2	PFALSAI	L40891 Plasmodium
19	838.2	56.6	843	2	PFALSAE	L40887 Plasmodium
20	838.2	56.6	843	2	PFALSAF	L40888 Plasmodium
21	838.2	56.6	843	2	PFALSAK	L40893 Plasmodium
22	838.2	56.6	843	2	PFALSAL	L40834 Plasmodium
23	836.6	56.5	843	2	PFALSAM	L40835 Plasmodium
24	835	56.3	843	2	PFALSA1N	L40947 Plasmodium
25	835	56.3	843	2	PFALSAD	L40886 Plasmodium
26	835	56.3	843	2	PFALSAN	L40836 Plasmodium
27	835	56.3	843	2	PFALSAO	L40837 Plasmodium
28	833.4	56.2	843	2	PFALSA1C	L40910 Plasmodium
29	833.4	56.2	843	2	PFALSAJ	L40892 Plasmodium
30	831.8	56.1	843	2	PFALSA1A	L40908 Plasmodium
31	831.8	56.1	843	2	PFALSA1B	L40909 Plasmodium
32	790.2	53.3	795	2	AF246996	AF246996 Plasmodiu
33	503.4	34.0	1368	11	AY751501	AY751501 Synthetic
34	472.8	31.9	988	2	PFLSA15	Z30319 P.falciparu
35	472.8	31.9	988	6	A28742	A28742 cDNA for LS
36	472.8	31.9	988	6	AR162963	AR162963 Sequence
37	472.8	31.9	988	6	AR178252	AR178252 Sequence
38	470.8	31.8	954	6	AR162966	AR162966 Sequence
39	470.8	31.8	954	6	AR178255	AR178255 Sequence
c 40	279	18.8	198224	9	AC149597	AC149597 Mus muscu
41	272.8	18.4	208332	9	AC122227	AC122227 Mus muscu
42	252.4	17.0	184015	9	AC144849	AC144849 Mus muscu
c 43	250.8	16.9	172716	14	AC115383	AC115383 Rattus no
c 44	235.8	15.9	161482	14	AC165336	AC165336 Mus muscu
45	224.2	15.1	2069	6	E10125	E10125 DNA encodin

ALIGNMENTS

RESULT 1
AR162974
LOCUS AR162974 1482 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 45 from patent US 6270771.
ACCESSION AR162974
VERSION AR162974.1 GI:16233440
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1482)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6270771-A 45 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1482
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1482; DB 6; Length 1482;
Best Local Similarity 100.0%; Pred. No. 9.8e-159;
Matches 1482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
```

```
Qy       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db       241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db       301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db       361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db       421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
Db       541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
Db       601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
Db       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db       841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
Db      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
Db      1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db      1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
Db      1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
Db      1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

RESULT 2
AR178263
LOCUS AR178263 1482 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 45 from patent US 6319502.
ACCESSION AR178263
VERSION AR178263.1 GI:20219401
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1482)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6319502-A 45 20-NOV-2001;
FEATURES Location/Qualifiers
source 1. .1482
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 1482; DB 6; Length 1482;
Best Local Similarity 100.0%; Pred. No. 9.8e-159;
Matches 1482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
```

```
Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

```
RESULT 3
AR162970
LOCUS       AR162970               1482 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 41 from patent US 6270771.
ACCESSION   AR162970
VERSION     AR162970.1  GI:16233435
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 41 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.9%;  Score 1480.4;  DB 6;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-158;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840
```

```
Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 4
AR178259
LOCUS       AR178259               1482 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 41 from patent US 6319502.
ACCESSION   AR178259
VERSION     AR178259.1  GI:20219397
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 41 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.9%;  Score 1480.4;  DB 6;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-158;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
```

```
Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
```

```
RESULT 5
PFLSA13
LOCUS       PFLSA13                 1496 bp    DNA     linear   INV 05-SEP-1994
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen-1 (3').
ACCESSION   Z30320
VERSION     Z30320.1  GI:510183
KEYWORDS    liver stage antigen-1; LSA-1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     The Plasmodium falciparum liver stage antigen LSA-1 is well
            conserved and harbors major B- and T-cell epitopes
  JOURNAL   J. Immunol. (1994) In press
REFERENCE   2  (bases 1 to 1496)
  AUTHORS   Fidock,D.A., Gras-Masse,H., Lepers,J., Brahimi,K., Benmohamed,L.,
            Mellouk,S., Guerin-Marchand,C., Londono,A., Raharimalala,L.,
            Meis,J.F., Langsley,G., Roussilhon,C., Tartar,A. and Druilhe,P.
  TITLE     Plasmodium falciparum liver stage antigen-1 is well conserved and
            contains potent B and T cell determinants
  JOURNAL   J. Immunol. 153 (1), 190-204 (1994)
   PUBMED   7515922
REFERENCE   3  (bases 1 to 1496)
  AUTHORS   Fidock,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1994) David A Fidock, Laboratory of bio-medical
            parasitology, Pasteur, Institute, 25, rue du Dr. Roux, PARIS, 75724
            CEDEX 15, France
FEATURES             Location/Qualifiers
     source          1. .1496
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="T9/96"
                     /db_xref="taxon:5833"
                     /clone="DG536, DG443, DG750"
                     /cell_type="human erythrocyte culture"
                     /clone_lib="Lambda gt11"
                     /dev_stage="asexual blood stage"
     gene            1. .1482
                     /gene="LSA-1"
     CDS             <1. .1482
                     /gene="LSA-1"
                     /note="nucl. 1 to 1496 corresponds to nucl. 4327 to 5822
                     of published NF54 strain
                     3' extremity of LSA-1 gene of strain T9/96"
                     /codon_start=1
                     /product="liver stage antigen-1"
                     /protein_id="CAA82975.1"
                     /db_xref="GI:510184"
                     /db_xref="UniProt/TrEMBL:Q25886"
                     /translation="QEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDRLAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDL
                     EQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQRDLEQRKADT
                     KKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYIPHQSSLPQDNRGNSRDSKE
                     ISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQKEDKSADIQNHTLETVNIS
                     DVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIVQYDNFQDEENIGIYKELED
                     LIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDNNFKPNDKSLYDEHIKKYKN
                     DKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKYFMKL"
     repeat_region   1. .639
ORIGIN
```

```
  Query Match             99.9%;  Score 1480.4;  DB 2;  Length 1496;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-158;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
```

```
Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 6
A28743
LOCUS       A28743                  1493 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA gene 3' (partial).
ACCESSION   A28743
VERSION     A28743.1  GI:1247516
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 32 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Query Match           97.5%;  Score 1444.4;  DB 6;  Length 1493;
Best Local Similarity 99.7%;  Pred. No. 1.7e-154;
Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
                |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 7
AR162967
LOCUS       AR162967               1493 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 38 from patent US 6270771.
ACCESSION   AR162967
VERSION     AR162967.1  GI:16233432
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 38 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             97.5%;  Score 1444.4;  DB 6;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-154;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
                 |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 8
AR178256
LOCUS       AR178256               1493 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 38 from patent US 6319502.
ACCESSION   AR178256
VERSION     AR178256.1  GI:20219394
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 38 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             97.5%;  Score 1444.4;  DB 6;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-154;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538
```

```
Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 9
PFLSA1G
LOCUS       PFLSA1G                5970 bp    DNA     linear   INV 18-APR-2005
DEFINITION  P.falciparum LSA-1 gene for liver stage antigen.
ACCESSION   X56203 S75010 S75012 S75014 S75016
VERSION     X56203.1  GI:9915
KEYWORDS    liver stage antigen.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
```

```
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Zhu,J. and Hollingdale,M.R.
  TITLE     Structure of Plasmodium falciparum liver stage antigen-1
  JOURNAL   Mol. Biochem. Parasitol. 48 (2), 223-226 (1991)
   PUBMED   1840628
REFERENCE   2  (bases 1 to 5970)
  AUTHORS   Zhu,J. and Hollingdale,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1990) Zhu J., Hollingdale M, Biomedical Research
            Institute, 12111 Parklawn Drive, Rockeville, MD 20852, U.S.A
COMMENT     On or before Jun 27, 2005 this sequence version replaced gi:241572,
            gi:241573, gi:241574, gi:241575.
FEATURES             Location/Qualifiers
     source          1. .5970
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="NF54"
                     /db_xref="taxon:5833"
                     /clone="RI-LSA segment, DRAI-LSA segment"
                     /cell_line="NF54 blood stage culture"
                     /clone_lib="pUC13 plasmid, RI-sublib, DRAI-sublib"
                     /dev_stage="liver stage"
     TATA_signal     30. .36
     gene            79. .5808
                     /gene="LSA-1"
     CDS             79. .5808
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage antigen"
                     /protein_id="CAA39663.1"
                     /db_xref="GI:9916"
                     /db_xref="HSSP:1D7M"
                     /db_xref="UniProt/TrEMBL:Q25893"
                     /translation="MKHILYISFYFILVNLLIFHINGKIIKNSEKDEIIKSNLRSGSS
                     NSRNRINEEKHEKKHVLSHNSYEKTKNNENNKFFDKDKELTMSNVKNVSQTNFKSLLR
                     NLGVSENIFLKENKLNKEGKLIEHIINDDDDKKKYIKGQDENRQEDLEEKAAKETLQG
                     QQSDLEQERLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQEQQSDLEQDRLA
                     KEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQE
                     RLAKEKLQEQQSDSEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKL
                     QEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDL
                     EQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLERTKASKETLQEQQSDLEQERLAK
                     EKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQ
                     SDLEQERRAKEKLQEQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQDR
                     LAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQ
                     EQQSDLEQDRLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLE
                     QERLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKE
                     KLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQS
                     DLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQGQQSDLEQERL
                     AKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQGQQSDLEQERLAKEKLQE
                     QQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLERTKASKETLQEQQSDLEQ
                     ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRAKEK
                     LQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD
                     LEQERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERRA
                     KEKLQEQQSDLEQDRLAKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQ
                     QSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDLEQE
                     RRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLANEKLQEQQRDLEQERRAKEKL
                     QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQRDL
                     EQERLAKEKLQEQQRDLEQRKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENE
                     RGYYIPHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGS
                     IKPEQKEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDE
                     FKPIVQYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYE
                     KTKDNNFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSE
                     DITKYFMKL"
ORIGIN

  Query Match           97.2%;  Score 1440.4;  DB 2;  Length 5970;
  Best Local Similarity 98.2%;  Pred. No. 2.7e-154;
```

```
  Matches 1456;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       4327 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 4386

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||||||||||||||||| |||| |||||||||||||||| ||||| || |||||||||
Db       4387 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 4446

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||| ||||||||||||||||||||||||||||||||||| ||||||||
Db       4447 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 4506

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||| |||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db       4507 GAGAGACTTGCTAAAGAAAAGTTGCAAGAGCAACAAAGAGATTTAGAACAAGAGAGACGT 4566

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
Db       4567 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 4626

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||| |||| || |||||||||||||||||||||||| |||||| ||||||||||||||
Db       4627 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAATGAAAAGTTGCAA 4686

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       4687 GAGCAACAAAGAGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 4746

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       4747 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 4806

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4807 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 4866

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              |||| |||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Db       4867 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGAGATTTAGAACAAGAGAGACTTGCT 4926

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4927 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 4986

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       4987 AATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTA 5046

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5047 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 5106

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       5107 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 5166

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5167 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 5226

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5227 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 5286

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5287 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 5346

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5347 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 5406
```

```
Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5407 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 5466

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5467 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 5526

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5527 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 5586

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5587 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 5646

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5647 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 5706

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5707 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 5766

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       5767 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 5808
```

```
RESULT 10
AE014834
LOCUS       AE014834               253001 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 10 section 6 of 7 of the
            complete sequence.
ACCESSION   AE014834 AE014185
VERSION     AE014834.1  GI:23495173
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253001)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253001)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253001
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
     repeat_region   complement(79. .107)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(113. .136)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   292. .441
                     /rpt_type=tandem
     repeat_region   complement(292. .441)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   455. .479
                     /rpt_type=tandem
     repeat_region   complement(455. .479)
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     gene            <498. .>3257
                     /locus_tag="PF10_0307"
     mRNA            <498. .>3257
                     /locus_tag="PF10_0307"
     CDS             498. .3257
                     /locus_tag="PF10_0307"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN35504.1"
                     /db_xref="GI:23495174"
                     /translation="MASCTGKVNFSSSGKEKGIGLESVLENAFDELISEFKKKFLHKI
                     IKVLEEAKYNENLDIMNERRKNFFLKKYENDQRDEDEYENKIGNVVEKKIEGNVEGN
                     VEGNVEREDIMIEEEKKKKKKKMDSSSSSNDSTFGVYLKSIIDRNKISLLKKLEKQKE
                     LKSVEENNKINYYNNKYKNGIIHYDENKEYHNNLNDKHVYNKNEESMNERTKLVKYTF
                     HDKNKNKNDMYLYPQESHFKYVSERIENLYKISNKLREEKETYHNLKYDKDNENQHFK
                     ENHTEDETDHEQNYEKYNVKLKNYVQNHNTQQMDKYQINEKNNISKLKQYNNNNDNNN
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLKGNMEKNIINNHQKNYISYYNNS
                     FKYPSRSHIFTERFSYHASDEHNEKSDHIYDMSPNKDKNCNNIYGILHNKDKHCNNNY
                     DMSSNKDKHCNNIYDMSPKKDKHCNNIYGILHNKDKHCNNIYGMSPKKDKHCNNIYGI
                     LHNKDKHCNNIYDMSPKKDKHCNNIYEMSPKKDKHCNNIYEMSPKKDKHCNNIYDNSH
                     LSSRGANKLPPSNNIIHEEKKQIINENINIESPQKINDLCNEKNNNLECKKEKIKLEK
                     QFSFDSKDSDDNYVRGKNNMYEQDDLKNNMSQRYYKKVEDFKSYQLNKKISNEKEIIK
                     RPNNNLFFEVIRGKKRKNMQGFECKDCKSFYNEIYSDDDEGKCVHKNFFMKNEQQLKK
                     KININKDVRSKNSSTDDINSVDKMNNVDKMNNANNINNADNINNADNINNADNINNED
                     NINNMDNLNNADNLNNADNLNNADNYPLNKKVHMNNSKKSIYYSHSLSPTCLNAHTDT
                     EKASMNEKKNGNMLNMDNIKKENDLEKEKEKKKKKIIQTFSRHRYHNKINDSPKNFWK
                     FDFFK"
     repeat_region   781. .821
                     /rpt_type=tandem
     repeat_region   841. .870
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   932. .1111
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   1156. .1175
                     /rpt_type=tandem
                     /rpt_unit="(taaaaa)n"
     repeat_region   complement(1369. .1397)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1439. .1601
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     repeat_region   1474. .1579
                     /rpt_type=tandem
     repeat_region   1479. .1588
                     /rpt_type=tandem
     repeat_region   complement(1618. .1669)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1756. .2184
                     /rpt_type=tandem
     repeat_region   1775. .1954
                     /rpt_type=tandem
                     /rpt_unit="(taa)n"
     repeat_region   complement(2707. .2735)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2772. .2814
                     /rpt_type=tandem
     repeat_region   2790. .2965
                     /rpt_type=tandem
```

```
repeat_region   2808. .2966
                /rpt_type=tandem
repeat_region   3142. .3171
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(3274. .3347)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3391. .3437)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(3439. .3503)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   3439. .3500
                /rpt_type=tandem
repeat_region   3440. .3500
                /rpt_type=tandem
repeat_region   complement(3561. .3683)
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(3710. .3745)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3763. .3824)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   3763. .3811
                /rpt_type=tandem
repeat_region   complement(3871. .4025)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3922. .3987)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   3926. .3987
                /rpt_type=tandem
repeat_region   complement(4037. .4082)
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   4043. .4082
                /rpt_type=tandem
repeat_region   4043. .4080
                /rpt_type=tandem
repeat_region   complement(4084. .4104)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(4135. .4181)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(4182. .4313)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   4364. .4466
                /rpt_type=tandem
repeat_region   4372. .4457
                /rpt_type=tandem
repeat_region   4377. .4472
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(4512. .4655)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4670. .4775)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   4675. .4768
                /rpt_type=tandem
repeat_region   4870. .4912
                /rpt_type=tandem
repeat_region   complement(4870. .4912)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
```

```
repeat_region   complement(4918. .5000)
                /rpt_type=tandem
                /rpt_unit="(taaaaa)n"
gene            <5162. .>8973
                /locus_tag="PF10_0308"
mRNA            join(<5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .>8973)
                /locus_tag="PF10_0308"
CDS             join(5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .8973)
                /locus_tag="PF10_0308"
                /codon_start=1
                /product="hypothetical protein"
```

```
  Query Match             94.8%;  Score 1405.2;  DB 2;  Length 253001;
  Best Local Similarity   96.8%;  Pred. No. 4.8e-151;
  Matches 1434;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              || |  |||||||||||||||||||||||||||  |||||||||||||| || |  |||
Db     179846 CAGGGGCAACAAAGCGATCTAGAACAAGAGAGACTAGCTAAAGAAAAGTTACAGGGGCAA 179905

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||||||| |||||||||| ||||| |||||||||||||||| | ||| |||||||||
Db     179906 CAAAGCGATCTAGAACAAGAGAGACTAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGAT 179965

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||||||||||||||||||| | |||||||||| ||||||||
Db     179966 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAGCGACAAAGCGATTTAGAACAA 180025

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180026 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTA 180085

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     180086 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGACAGACTTGCTAAAGAA 180145

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||| ||||| || |||||||||||||||||||||||||  ||||| |||||||| ||
Db     180146 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTAGCTAAGGAAAAGTTACAG 180205

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              | || || ||||||||||||| |  ||       |||||||| ||||||||||||||||
Db     180206 GGGCAGCATAGCGATTTAGAACGAACGAAGGCATCTAAAGAAACGTTGCAAGAACAACAA 180265

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180266 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 180325

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180326 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 180385

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180386 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 180445

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180446 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 180505

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     180506 AATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTA 180565

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180566 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 180625

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db     180626 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCGAAGGAAATATCTATAATA 180685
```

```
Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180686 GAAAATACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 180745

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180746 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 180805

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180806 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 180865

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180866 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 180925

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180926 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 180985

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180986 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 181045

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181046 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 181105

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181106 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 181165

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181166 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 181225

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181226 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 181285

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db     181286 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 181327


RESULT 11
A28740
LOCUS       A28740                   950 bp    DNA     linear   PAT 04-JUN-1995
DEFINITION  cDNA for LSA-R-NR protein.
ACCESSION   A28740
VERSION     A28740.1  GI:1247512
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   .
  TITLE     LIVER-STAGE-SPECIFIC PEPTIDE SEQUENCES OF P. FALCIPARUM BEARING
            EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES
  JOURNAL   Patent: WO 9213884-A 29 20-AUG-1992;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             64.0%;  Score 948.4;  DB 6;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 2.3e-98;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 12
AR162961
LOCUS       AR162961                 950 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 32 from patent US 6270771.
ACCESSION   AR162961
VERSION     AR162961.1  GI:16233425
KEYWORDS    .
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 32 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           64.0%;  Score 948.4;  DB 6;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 2.3e-98;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 13
AR178250
LOCUS       AR178250                950 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 32 from patent US 6319502.
ACCESSION   AR178250
VERSION     AR178250.1  GI:20219388
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 32 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           64.0%;  Score 948.4;  DB 6;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 2.3e-98;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950
```

RESULT 14
PFALSAC

```
LOCUS       PFALSAC                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clones KEN1-21b, KEN1-40h, KEN1-56b,
            PNG1-28a, PNG1-28c, BRA1-19e) liver stage-specific antigen 1
            (LSA-1) gene, 3' end of cds.
ACCESSION   L40885
VERSION     L40885.1  GI:725278
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN1-56b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG1-28a) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG1-28c) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: KEN1-21b) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN1-40h) blood stage parasite DNA; and Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA1-19e) blood
            stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN1-56b"
                     /clone="PNG1-28a"
                     /clone="PNG1-28c"
                     /clone="KEN1-21b"
                     /clone="KEN1-40h"
                     /clone="BRA1-19e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59231.1"
                     /db_xref="GI:950664"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       228
                     /gene="LSA-1"
                     /note="clone PNG1-28a"
                     /replace="t"
ORIGIN
```

```
  Query Match             56.8%;  Score 841.4;  DB 2;  Length 843;
  Best Local Similarity   99.9%;  Pred. No. 3e-86;
  Matches  842;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

RESULT 15
PFALSAB

```
LOCUS       PFALSAB                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium flaciparum (clones KEN0-11a,-11b,-37d,-39a,-44e,-51a,
            -53c,-53d, -56a, PNG0-5h,-7a,-8f,-19a,-26b,-86h,-86i, BRA0-5a)
            liver stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40884
VERSION     L40884.1  GI:725277
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-44e) DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: KEN0-51a) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            PNG0-7a) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG0-5h) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            KEN0-11a) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: KEN0-11b) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            KEN0-37d) DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: KEN0-39a) DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-8f) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-19a) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-26b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-26b) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG0-86h) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG0-86i) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-53c) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: KEN0-53d) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN0-56a) blood stage parasite DNA; and Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA0-5a) blood stage
            parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN0-51a"
                     /clone="KEN0-44e"
                     /clone="PNG0-7a"
                     /clone="PNG0-5h"
                     /clone="KEN0-11a"
                     /clone="KEN0-11b"
                     /clone="KEN0-37d"
                     /clone="KEN0-39a"
                     /clone="PNG0-8f"
                     /clone="PNG0-19a"
                     /clone="PNG0-26b"
                     /clone="PNG0-86h"
                     /clone="PNG0-86i"
                     /clone="KEN0-53c"
                     /clone="KEN0-53d"
                     /clone="KEN0-56a"
                     /clone="BRA0-5a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59230.1"
                     /db_xref="GI:950663"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       60
                     /gene="LSA-1"
                     /note="clone BRA0-5a"
                     /replace="g"
ORIGIN
```

```
  Query Match             56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 4.5e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179

Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239

Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299

Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359

Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419

Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479

Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482

Db        841 TAA 843
```

```
RESULT 16
PFALSAG
LOCUS       PFALSAG                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN5-44d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40889
VERSION     L40889.1  GI:725282
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN5-44d) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
            Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN5-44d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41599.1"
                     /db_xref="GI:950668"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPKQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
  Query Match           56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 4.5e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699

Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819

Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879

Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939

Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999

Db        301 AAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059

Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119

Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179

Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239

Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299

Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359

Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419

Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479

Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482

Db        841 TAA 843
```

```
RESULT 17
PFALSAH
LOCUS       PFALSAH                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN6-21a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40890
VERSION     L40890.1  GI:725283
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN6-21a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN6-21a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41600.1"
                     /db_xref="GI:950669"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEYIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 4.5e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGTATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 18
PFALSAI
LOCUS       PFALSAI                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN7-40c) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40891
VERSION     L40891.1  GI:725284
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN7-40c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates 1.
```

```
           Last author is Altaf A. Lal.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN7-40c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41601.1"
                     /db_xref="GI:950670"
                     /translation="RKADTKKNLERKKEHGDILAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDY
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match             56.7%;  Score 839.8;  DB 2;  Length 843;
  Best Local Similarity   99.8%;  Pred. No. 4.5e-86;
  Matches  841;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATTATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 19
PFALSAE
LOCUS       PFALSAE                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN3-39d) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40887
VERSION     L40887.1  GI:725280
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN3-39d) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN3-39d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41597.1"
                     /db_xref="GI:950666"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDASLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match             56.6%;  Score 838.2;  DB 2;  Length 843;
  Best Local Similarity   99.6%;  Pred. No. 6.9e-86;
  Matches  840;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCTTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 20
PFALSAF
LOCUS       PFALSAF                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN4-32a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40888
VERSION     L40888.1  GI:725281
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
  PUBMED    7477115
COMMENT     Original source text: Plasmodium falciparum (clone: KEN4-32a) blood
            stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /clone="KEN4-32a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41598.1"
                     /db_xref="GI:950667"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKMKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.6%;  Score 838.2;  DB 2;  Length 843;
  Best Local Similarity  99.6%;  Pred. No. 6.9e-86;
  Matches  840;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```


```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATGAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 21
PFALSAK
LOCUS       PFALSAK                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN9-37e) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40893
VERSION     L40893.1  GI:725286
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN9-37e) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN9-37e"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41603.1"
                     /db_xref="GI:950672"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKDENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.6%;  Score 838.2;  DB 2;  Length 843;
  Best Local Similarity 99.6%;  Pred. No. 6.9e-86;
  Matches  840;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAGATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 22
PFALSAL
LOCUS       PFALSAL                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone PNG3-5f,-68f,-68h) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40834
VERSION     L40834.1  GI:725287
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-5f) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG3-68h) blood
            stage parasite DNA; and Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG3-68f) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="PNG3-5f"
                     /clone="PNG3-68h"
                     /clone="PNG3-68f"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59233.1"
                     /db_xref="GI:950673"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPAIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEE
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNDKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       573
                     /gene="LSA-1"
                     /note="clone PNG3-68h"
                     /replace="c"
ORIGIN

  Query Match           56.6%;  Score 838.2;  DB 2;  Length 843;
  Best Local Similarity   99.6%;  Pred. No. 6.9e-86;
  Matches  840;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAGAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843

RESULT 23
PFALSAM
LOCUS       PFALSAM                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clones BRA3-5b,-19a,-14b,-28e) liver
            stage-specific antigen 1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40835
VERSION     L40835.1  GI:725288
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
```

```
  PUBMED   7477115
COMMENT    Original source text: Plasmodium falciparum (individual_isolate
           Kenyan) (clone: BRA3-14b) blood stage parasite DNA; Plasmodium
           falciparum (individual_isolate Kenyan) (clone: BRA3-5b) blood stage
           parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
           (clone: BRA3-28e) blood stage parasite DNA; Plasmodium falciparum
           (individual_isolate Kenyan) (clone: BRA3-29b) blood stage parasite
           DNA; and Plasmodium falciparum (individual_isolate Kenyan) (clone:
           BRA3-19a) blood stage parasite DNA.
           1. Last author is Altaf A. Lal 2. The following clones have the sam
           e nucleotide sequence: BRA3-5b; BRA3-19a; BRA3-14b (one silent
           change at 5181,
           A to G) and BRA3-28e (one silent change at 5040, T to C) Clone name
           abbreviations: KEN for Kenya; PNG for Papua New Guinea and BRA for
           Brazil. For each clone, the number following the KEN, PNG or BRA
           represents group number, the number following the '-' denotes the
           patient's blood sample number and the lower case letter represents
           recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="BRA3-14b"
                     /clone="BRA3-5b"
                     /clone="BRA3-28e"
                     /clone="BRA3-29b"
                     /clone="BRA3-19a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59234.1"
                     /db_xref="GI:950674"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       75
                     /gene="LSA-1"
                     /note="clone BRA3-28e"
                     /replace="c"
     variation       216
                     /gene="LSA-1"
                     /note="clone BRA3-14b"
                     /replace="g"
ORIGIN

  Query Match           56.5%;  Score 836.6;  DB 2;  Length 843;
  Best Local Similarity   99.5%;  Pred. No. 1e-85;
  Matches  839;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843
```

```
RESULT 24
PFALSA1N
LOCUS       PFALSA1N                   843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40947
VERSION     L40947.1  GI:725317
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
  PUBMED    7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA6-29c) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
```

```
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA6-29c"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41592.1"
                     /db_xref="GI:725318"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVSDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match          56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity  99.4%;  Pred. No. 1.6e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAGTGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 25
PFALSAD
LOCUS       PFALSAD                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN2-51b, PNG2-7f,-9f,-11a,-11b,-19d,
            -26e, BRA2-14c,-30b,-30) liver stage-specific antigen 1 (LSA-1)
            gene, 3' end of cds.
ACCESSION   L40886
VERSION     L40886.1  GI:725279
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN2-51b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-30e) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: PNG2-26e) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: BRA2-14c) blood
            stage parasite DNA; Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA2-30b) blood stage parasite DNA; Plasmodium
            falciparum (individual_isolate Kenyan) (clone: PNG2-7f) blood stage
            parasite DNA; Plasmodium falciparum (individual_isolate Kenyan)
            (clone: PNG2-9f) blood stage parasite DNA; Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-11a) blood stage parasite
            DNA; Plasmodium falciparum (individual_isolate Kenyan) (clone:
            PNG2-11b) blood stage parasite DNA; and Plasmodium falciparum
            (individual_isolate Kenyan) (clone: PNG2-19d) blood stage parasite
            DNA.
            1. Last author is Altaf A. Lal 2. Following clones also have the sa
            me nucleotide sequence: KEN121b, KEN140h, KEN156b, PNG1281 (at 5193
            A to T), P
            NG128c, BRA119e
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN2-51b"
                     /clone="BRA2-30e"
                     /clone="PNG2-26e"
                     /clone="BRA2-14c"
                     /clone="BRA2-30b"
```

```
                     /clone="PNG2-7f"
                     /clone="PNG2-9f"
                     /clone="PNG2-11a"
                     /clone="PNG2-11b"
                     /clone="PNG2-19d"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59232.1"
                     /db_xref="GI:950665"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       222
                     /gene="LSA-1"
                     /note="clone PNG2-9f"
                     /replace="c"
     variation       240
                     /gene="LSA-1"
                     /note="clones PNG2-7f, BRA2-30b and BRA2-30e"
                     /replace="g"
ORIGIN

  Query Match             56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity   99.4%;  Pred. No. 1.6e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 26
PFALSAN
LOCUS       PFALSAN                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA4-28a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40836
VERSION     L40836.1  GI:725289
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: BRA4-28a) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="BRA4-28a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41604.1"
                     /db_xref="GI:950675"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEERIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
  Query Match           56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity  99.4%;  Pred. No. 1.6e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAGAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 27
PFALSAO
LOCUS       PFALSAO                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA5-32f) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40837
VERSION     L40837.1  GI:725290
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (clone: BRA5-32f) blood
            stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /clone="BRA5-32f"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41605.1"
                     /db_xref="GI:950676"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEISAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEENIGIYKELEDLIEKNGNLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.3%;  Score 835;  DB 2;  Length 843;
  Best Local Similarity  99.4%;  Pred. No. 1.6e-85;
  Matches  838;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGGAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
                |||
Db        841 TAA 843
```

```
RESULT 28
PFALSA1C
LOCUS       PFALSA1C                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA9-32b) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40910
VERSION     L40910.1  GI:725297
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA9-32b) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA9-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41582.1"
                     /db_xref="GI:725298"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN
```

```
  Query Match             56.2%;  Score 833.4;  DB 2;  Length 843;
  Best Local Similarity   99.3%;  Pred. No. 2.4e-85;
  Matches  837;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAGCATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840
```

```
Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 29
PFALSAJ
LOCUS       PFALSAJ                  843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone KEN8-32b) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40892
VERSION     L40892.1  GI:725285
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Kenyan) (clone: KEN8-32b) blood stage parasite DNA.
            1. Last author is Altaf A. Lal
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
            and BRA for Brazil. For each clone, the number following the KEN,
            PNG or BRA represents group number, the number following the '-'
            denotes the patient's blood sample number and the lower case letter
            represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Kenyan"
                     /db_xref="taxon:5833"
                     /clone="KEN8-32b"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAC41602.1"
                     /db_xref="GI:950671"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFKPIV
                     QYDNFQDEESIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
ORIGIN

  Query Match           56.2%;  Score 833.4;  DB 2;  Length 843;
  Best Local Similarity   99.3%;  Pred. No. 2.4e-85;
  Matches  837;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAGCATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


RESULT 30
PFALSA1A
LOCUS       PFALSA1A                 843 bp    DNA     linear   INV 19-AUG-1995
DEFINITION  Plasmodium falciparum (clone BRA7-30a) liver stage-specific antigen
            1 (LSA-1) gene, 3' end of cds.
ACCESSION   L40908
VERSION     L40908.1  GI:725293
KEYWORDS    exoerythrocytic; liver stage-specific antigen; liver stage-specific
            antigen 1; malaria.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Yang,C., Shi,Y.P., Udhayakumar,V., Alpers,M.P., Povoa,M.M.,
            Hawley,W.A., Collins,W.E. and Lal,A.A.
  TITLE     Sequence variations in the non-repetitive regions of the liver
            stage-specific antigen-1 (LSA-1) of Plasmodium falciparum from
            field isolates
  JOURNAL   Mol. Biochem. Parasitol. 71 (2), 291-294 (1995)
   PUBMED   7477115
COMMENT     Original source text: Plasmodium falciparum (individual_isolate
            Brazilian) (clone: BRA7-30a) blood stage parasite DNA.
            Clone name abbreviations: KEN for Kenya; PNG for Papua New Guinea
```

```
          and BRA for Brazil. For each clone, the number following the KEN,
          PNG or BRA represents group number, the number following the '-'
          denotes the patient's blood sample number and the lower case letter
          represents recombinant clones from the corresponding isolates.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="Brazilian"
                     /db_xref="taxon:5833"
                     /clone="BRA7-30a"
                     /dev_stage="blood stage parasite"
     gene            1. .843
                     /gene="LSA-1"
     CDS             <1. .843
                     /gene="LSA-1"
                     /codon_start=1
                     /product="liver stage-specific antigen 1"
                     /protein_id="AAB59228.1"
                     /db_xref="GI:725294"
                     /translation="RKADTKKNLERKKEHGDVLAEDLYGRLEIPVIELPSENERGYYI
                     PHQSSLPQDNRGNSRDSKEISIIEKTNRESITTNVEGRRDIHKGHLEEKKDGSIKPEQ
                     KEDKSADIQNHTLETVNISDVNDFQISKYEDEIGAEYDDSLIDEEEDDEDLDEFRPIV
                     QYDNFQDEENIGIYKELEDLIEKNENLDDLDEGIEKSSEELSEEKIKKGKKYEKTKDN
                     NFKPNDKSLYDEHIKKYKNVKQVNKEKEKFIKSLFHIFDGDNEILQIVDELSEDITKY
                     FMKL"
     variation       240
                     /gene="LSA-1"
                     /note="clone BRA7-30a"
                     /replace="g"
ORIGIN

  Query Match             56.1%;  Score 831.8;  DB 2;  Length 843;
  Best Local Similarity   99.2%;  Pred. No. 3.6e-85;
  Matches  836;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATGTATTAGCA 60

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 GAGGATTTATATGGTCGTTTAGAAATACCAGTTATAGAACTTCCATCAGAAAATGAACGT 120

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 180

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCAAGGAAATATCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAG 240

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 300

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 360

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        361 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAGGTGCTGAATATGAC 420

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        421 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAGGCCTATTGTG 480

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 540

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 600

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 660

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        661 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGTTAAGCAG 720

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 780

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 840

Qy       1480 TAA 1482
              |||
Db        841 TAA 843


Search completed: December 20, 2005, 10:43:12
Job time : 8595.96 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:12:49 ; Search time 331.317 Seconds
(without alignments)
7951.129 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1482	100.0	1482	3	US-08-098-327E-45	Sequence 45, Appl
	2	1482	100.0	1482	3	US-08-462-625-45	Sequence 45, Appl
	3	1480.4	99.9	1482	3	US-08-098-327E-41	Sequence 41, Appl
	4	1480.4	99.9	1482	3	US-08-462-625-41	Sequence 41, Appl
	5	1444.4	97.5	1493	3	US-08-098-327E-38	Sequence 38, Appl
	6	1444.4	97.5	1493	3	US-08-462-625-38	Sequence 38, Appl
	7	948.4	64.0	950	3	US-08-098-327E-32	Sequence 32, Appl
	8	948.4	64.0	950	3	US-08-462-625-32	Sequence 32, Appl
	9	851.6	57.5	1320	2	US-08-257-073-15	Sequence 15, Appl
	10	472.8	31.9	988	3	US-08-098-327E-34	Sequence 34, Appl
	11	472.8	31.9	988	3	US-08-462-625-34	Sequence 34, Appl
	12	470.8	31.8	954	3	US-08-098-327E-37	Sequence 37, Appl
	13	470.8	31.8	954	3	US-08-462-625-37	Sequence 37, Appl
	14	128.2	8.7	1039	3	US-09-902-540-1280	Sequence 1280, Ap
	15	119.2	8.0	612	3	US-09-902-540-1357	Sequence 1357, Ap
	16	114.6	7.7	3095	9	5231168-1	Patent No. 5231168
	17	111.4	7.5	1827	3	US-09-270-767-1308	Sequence 1308, Ap
	18	111.4	7.5	1827	3	US-09-270-767-16590	Sequence 16590, A
c	19	110.2	7.4	7218	2	US-08-232-463-14	Sequence 14, Appl
	20	99.6	6.7	1956	3	US-08-559-896B-1	Sequence 1, Appli
	21	99.6	6.7	1956	3	US-09-351-794A-1	Sequence 1, Appli
	22	93.6	6.3	614	3	US-09-902-540-1318	Sequence 1318, Ap
	23	93.4	6.3	187169	3	US-09-949-016-12776	Sequence 12776, A
	24	93.4	6.3	191569	3	US-09-949-016-15940	Sequence 15940, A
	25	92.6	6.2	5361	3	US-08-973-462-2	Sequence 2, Appli
	26	92.6	6.2	6152	3	US-08-973-462-1	Sequence 1, Appli
c	27	91.6	6.2	29927	3	US-09-949-016-11814	Sequence 11814, A
c	28	91.6	6.2	29927	3	US-09-949-016-17474	Sequence 17474, A
c	29	91.6	6.2	29927	3	US-09-949-016-17475	Sequence 17475, A
c	30	91.6	6.2	37802	3	US-09-949-016-12639	Sequence 12639, A
	31	88.4	6.0	4766	6	PCT-US93-07261-10	Sequence 10, Appl
c	32	88.4	6.0	5340	3	US-09-627-122-21	Sequence 21, Appl
	33	88	5.9	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	34	88	5.9	1664976	3	US-09-692-570-1	Sequence 1, Appli
	35	87	5.9	699	3	US-09-248-796A-9722	Sequence 9722, Ap
	36	86.4	5.8	601	3	US-09-949-016-203127	Sequence 203127,
	37	86.4	5.8	601	3	US-09-949-016-203169	Sequence 203169,
	38	84.8	5.7	1669	3	US-09-461-697-184	Sequence 184, App
c	39	83.4	5.6	147382	3	US-09-949-016-14624	Sequence 14624, A
	40	82.8	5.6	696	3	US-09-461-697-193	Sequence 193, App
	41	82.8	5.6	699	3	US-09-461-697-191	Sequence 191, App
	42	82.8	5.6	717	3	US-09-461-697-189	Sequence 189, App
	43	82.8	5.6	774	3	US-09-461-697-187	Sequence 187, App
	44	82.8	5.6	819	3	US-09-461-697-185	Sequence 185, App
	45	82.8	5.6	19124	2	US-08-487-826B-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT 1
US-08-098-327E-45
; Sequence 45, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
```

```
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-45

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-247;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 2
US-08-462-625-45
; Sequence 45, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
```

```
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-45

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-247;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 3
US-08-098-327E-41
; Sequence 41, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1482
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-41

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 2.5e-247;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 4
US-08-462-625-41
; Sequence 41, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Burns, Doane, Swecker & Mathis
;       STREET:  P.O. Box 1404
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  United States
;       ZIP:  22313-1404
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/462,625
;       FILING DATE:  05-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/098,327
;       FILING DATE:  24-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 91 01286
;       FILING DATE:  05-FEB-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McGowan, Malcolm K.
;       REGISTRATION NUMBER:  39,300
;       REFERENCE/DOCKET NUMBER:  010830-078
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  41:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1482 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1482
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  WO 92/13884
;       PUBLICATION DATE:  20-AUG-1992
US-08-462-625-41

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 2.5e-247;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
```

```
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 5
US-08-098-327E-38
; Sequence 38, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-38

  Query Match             97.5%;  Score 1444.4;  DB 3;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 4.1e-241;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 6
US-08-462-625-38
; Sequence 38, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-38

  Query Match             97.5%;  Score 1444.4;  DB 3;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 4.1e-241;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197
```

```
Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479
```

RESULT 7
US-08-098-327E-32
; Sequence 32, Application US/08098327E
; Patent No. 6270771
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine
; APPLICANT: DRUILHE, Pierre
; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/098,327E
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 01286
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 92/13884
; PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-32

Query Match 64.0%; Score 948.4; DB 3; Length 950;
Best Local Similarity 99.9%; Pred. No. 2.1e-155;

Matches 949; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950
```

RESULT 8
US-08-462-625-32
; Sequence 32, Application US/08462625
; Patent No. 6319502
; GENERAL INFORMATION:
; APPLICANT: GUERIN-MARCHAND, Claudine

```
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-32

  Query Match             64.0%;  Score 948.4;  DB 3;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-155;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy         611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy         671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy         731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy         791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy         851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy         911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 9
US-08-257-073-15
; Sequence 15, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
```

```
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-15

  Query Match             57.5%;  Score 851.6;  DB 2;  Length 1320;
  Best Local Similarity   87.8%;  Pred. No. 1.1e-138;
  Matches  929;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        425 ATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 484
              || || || | | |  ||    | ||     ||     |    | | |     | | |   |
Db        263 ATGTAAAAAATGTGTCACAAACAAATTTCAAAAGTCTTTTAAGAAATCTTGGTGTTTCAG 322

Qy        485 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 544
              |  | | |     |||||||||       | |  |  || |  | |||     || | | |
Db        323 AGAATATATTCCTTAAAGAAAATAAATTAAATAAGGAAGGGAAATTAATTGAACACATAA 382

Qy        545 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAG 604
                | | ||  |     || || || | |   |     || | || |   |     |||
Db        383 TAAATGATGATGACGATAAAAAAAAATATATTAAAGGGCAAGACGAAAACAGACAAGAAG 442

Qy        605 AAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATT 664
              |          || |   ||     |  |  ||||   ||||||||||||||||||||||
Db        443 ATCTTGAAGAAAAAGCGCGCGCATCTAAAGAAACGAGGAAGGCTGATACGAAAAAAAATT 502

Qy        665 TAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAA 724
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        503 TAGAAAGAAAAAAGGAACATGGAGATGTATTAGCAGAGGATTTATATGGTCGTTTAGAAA 562

Qy        725 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 622

Qy        785 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAA 844
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        623 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAA 682

Qy        845 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 742

Qy        905 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 802

Qy        965 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 1024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 862

Qy       1025 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 1084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 922

Qy       1085 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 1144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 982

Qy       1145 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1042

Qy       1205 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1102

Qy       1265 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1162

Qy       1325 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1163 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1222

Qy       1385 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGT 1282

Qy       1445 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||
Db       1283 TATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1320


RESULT 10
US-08-098-327E-34
; Sequence 34, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-34
```

```
  Query Match           31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity  95.7%;  Pred. No. 3.2e-73;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              ||||  |  |  |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db        661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 11
US-08-462-625-34
; Sequence 34, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  988 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-34

  Query Match           31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity  95.7%;  Pred. No. 3.2e-73;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              ||||  |  |  |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db        661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
              ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 12
US-08-098-327E-37
; Sequence 37, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 7.1e-73;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db        509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              |||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db        629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db        809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

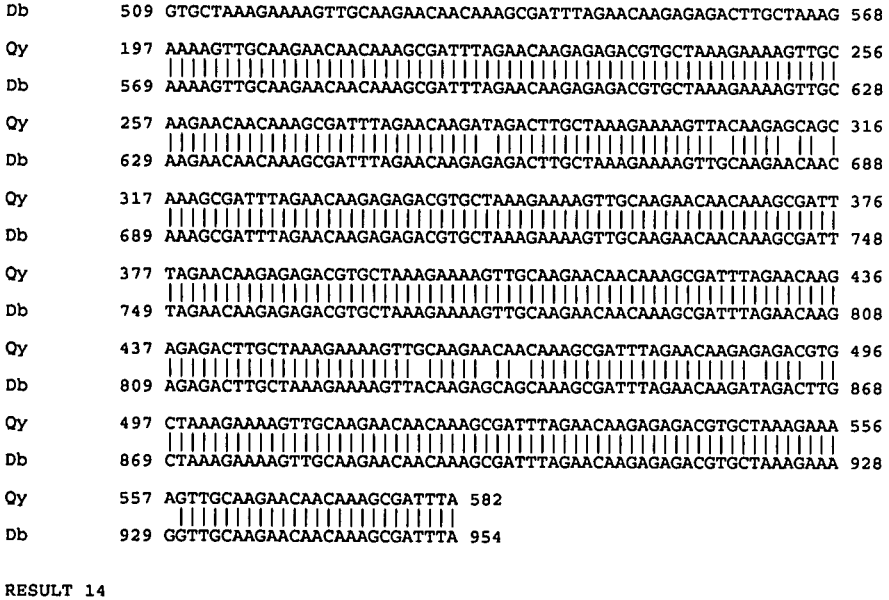
Qy        557 AGTTGCAAGAACAACAAAGCGATTTA 582
              |||||||||||||||||||||||||
Db        929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 13
US-08-462-625-37
; Sequence 37, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 7.1e-73;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

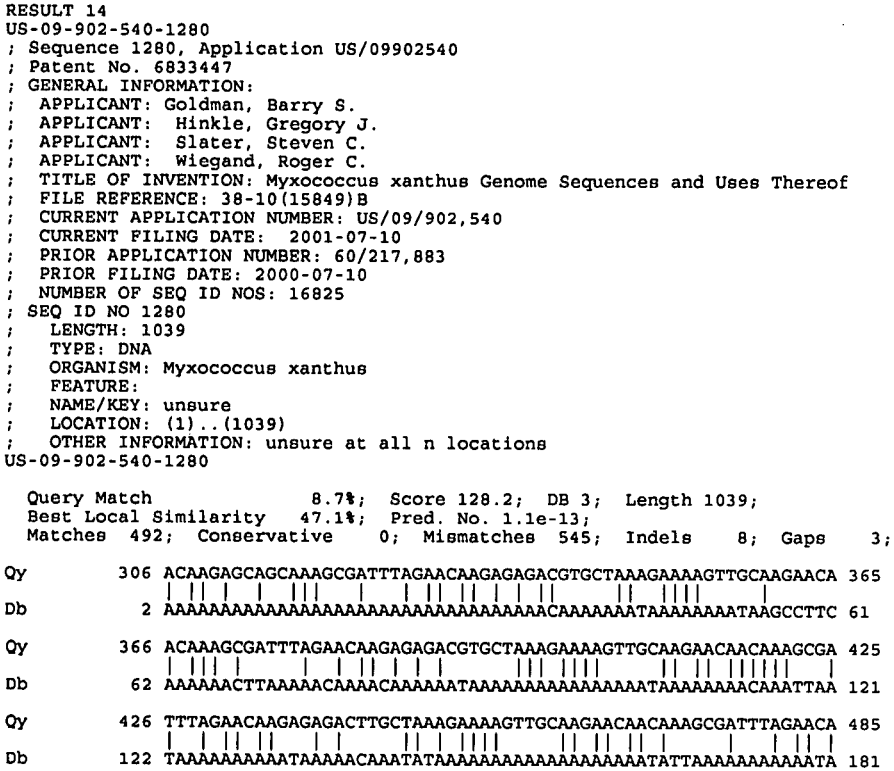
Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

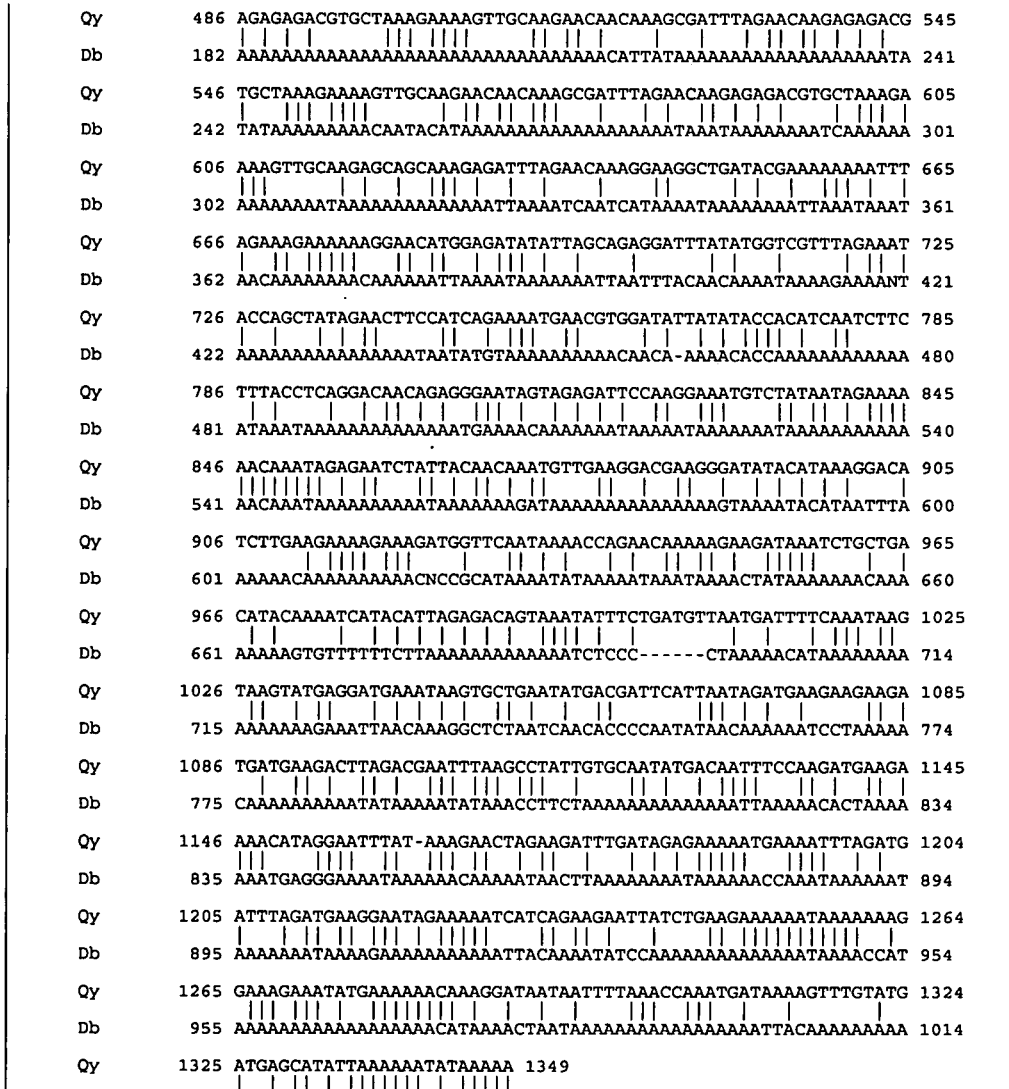
Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
```



RESULT 14
US-09-902-540-1280
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 8.7%; Score 128.2; DB 3; Length 1039;
Best Local Similarity 47.1%; Pred. No. 1.1e-13;
Matches 492; Conservative 0; Mismatches 545; Indels 8; Gaps 3;





RESULT 15
US-09-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540

```
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(612)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

  Query Match             8.0%;  Score 119.2;  DB 3;  Length 612;
  Best Local Similarity   50.2%;  Pred. No. 3.9e-12;
  Matches  292;  Conservative    0;  Mismatches  290;  Indels    0;  Gaps    0;

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
              || |  | | |     ||  |||      | || | |||||| ||| | | ||   |  | |
Db          9 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 68

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
              | |||      ||  |  ||    | | | | |||  | |      || |||| ||| |
Db         69 AAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 128

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
                || ||||  | ||  |||||  | | |||       | | | |||||  | | |  | |
Db        129 AAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 188

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
              | |  ||| |       |     ||  |      ||| ||  ||  |  |  |  ||| ||
Db        189 AAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAATA 248

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                || |  |  |    |  |||| |  ||  | || || |  || |  |  |  ||
Db        249 AAAAAAAAAAATAAAATAAAAATAAAAAAAAATAAAATATTAAAAAAAAATTAAAAAAAAA 308

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
              ||   ||   |  || |  | ||     || |  |  |||| | |  || | | ||| ||
Db        309 AAAANTAAAATAAAAAAAAAAAAAAAANTAAAAAAATAAAAAAAAAAATAAATAAAAAAAAA 368

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              || || |    | | | ||  |||    | |   |  | || ||  || | |||||
Db        369 ATAAAAAAAAAAAAATAAAATATAAAAATAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAA 428

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                | | || | | |    ||   |||| | || ||  ||| ||| || |||||| |||
Db        429 AAAAAAAAAAATTAAAAAAAAATTAAAAAATAATAAAAAAAAAAAAAATAAAAAAAAAAAA 488

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
              | || ||    |||  |||  | ||||      |  |  |  | |  |||||| | |||
Db        489 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 548

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              ||  || |  |    || ||  ||||  |||||   | ||||
Db        549 AAAAATTAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 590


RESULT 16
5231168-1
;Patent No. 5231168
;    APPLICANT: DZIEGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN;
;VUUST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
;    TITLE OF INVENTION: MALARIA ANTIGEN
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/409,658
;      FILING DATE: 18-SEP-1989
;SEQ ID NO:1:
;      LENGTH: 3095
5231168-1

  Query Match             7.7%;  Score 114.6;  DB 9;  Length 3095;
  Best Local Similarity   44.3%;  Pred. No. 2.8e-11;
  Matches  650;  Conservative    0;  Mismatches  804;  Indels   13;  Gaps    4;

Qy         17 ATCTAGAACAAGAGAGACGTGCTAAAGAAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 76
              || |  | ||  || | |   |    ||||  |  | |    | ||| |    |   |||
Db       1133 ATATTCAGCATGAAATAGTAGAAGTAGAAGAAATACTTCCAGAAGATGATAAAAATGAAA 1192

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
              ||| |  || ||  | || | |  |  ||||    | |   ||     || | || | |
Db       1193 AAGTTGAACATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAG 1252

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
              |    | ||||   |  | |   || |||      | |||| | || | |  || ||||
Db       1253 GTCAACATGAAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATGATAAAAATGAAAAAG 1312

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
              |     || |  | ||   ||  |   | |     |  |  || |||||    |
Db       1313 TTGAACATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTC 1372

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
              || |  ||  |   ||  | ||| ||  |  ||  |   |  ||||   | |||   | |
Db       1373 AACATGAAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTGAAC 1432

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
              |         |||||   ||   |  |    |||| |      | ||| ||       ||
Db       1433 ATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTCAACATG 1492

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
              |  | ||||      | | ||||||   | | |||| |  |||  ||  ||  |||
Db       1493 AAATAGTAGAG------GTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAAC 1546

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
              |   | | |   |||   |        || | ||    |   ||| |||    || ||
Db       1547 ATGAAATAGTAGAAGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGGTCAACATG 1606

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
              | || | || |  |||||   | |   ||   |||  || |   |   | | || | |
Db       1607 AAATAGTAGAGGTTGAAGAAATTCTACCAGAAGAAGATAAAAATGAAAAAGGTCAACATG 1666

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 616
              | |   |||     ||      | ||  | |||| | |  ||  ||||   |     ||
Db       1667 AAATAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAACATGAAA 1726

Qy        617 AGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAA 676
              |     ||  |  | |    | | | | |  ||||| |  | |  || |
Db       1727 TAGTAGAGGTTGAAGAAATTCTACCAGAAGATAAAAATGAAAAAGTTCAACATGAAATAG 1786

Qy        677 AGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAG 736
              ||   || ||| ||  |  ||||   | |  |    |  || | | ||   | |
Db       1787 TAGAGGTTGAAGAAATTCTTCCAGAAATTGTTGAAATTGAAGAAGTACCATCACAAACAA 1846

Qy        737 AACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGG 796
              |          | | | ||||  |  | |       | | |  |   |  ||   |   |
Db       1847 ATAACAATGAAAATATTGAAACTATAAAACCAGAAGAAAAAAAGAATGAATTTAGTGTTG 1906

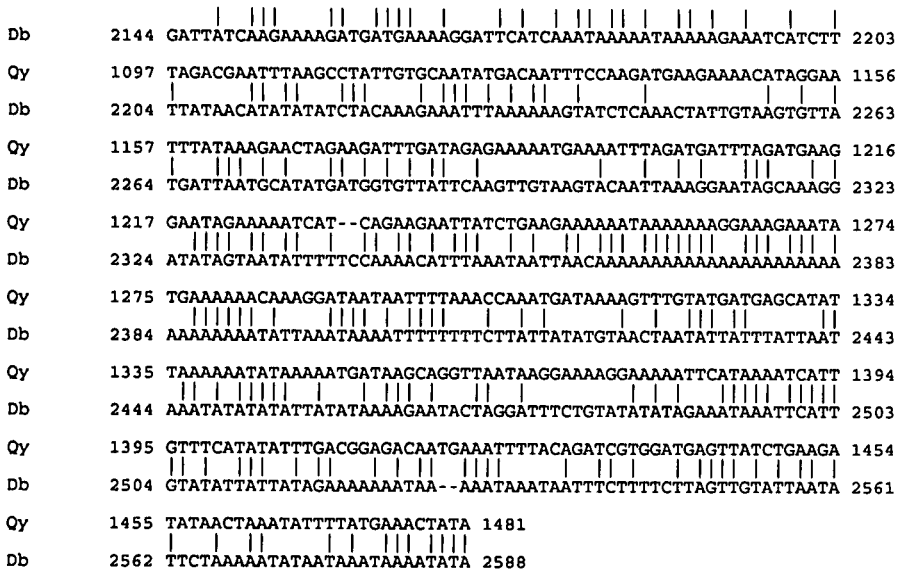
Qy        797 ACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAG 856
              |  | | ||  | |     |||  ||  || |    ||| | || |  |||   | |  |
Db       1907 AAGAAAAAGCAATTCCACAAGAACCCGTGGTAC---CTACATTAAATGAAAATGAAAACG 1963

Qy        857 AATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAA 916
              ||   | | ||  ||  |  |   |    | ||| ||  ||  || ||  |   |
Db       1964 TTACTCCCAAACCATCTGAAGGTGAATCCACTAAACCAGATATAGTTCAAATTAAAATAG 2023

Qy        917 AGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATC 976
              ||||   |  |  |||   | |  ||| |  ||    |  || |   | |||
Db       2024 TACAAGAAAATAAACCAAATAAAAAGGAAACACCAGTAGTAGATGGTCCAAAACATGTAG 2083

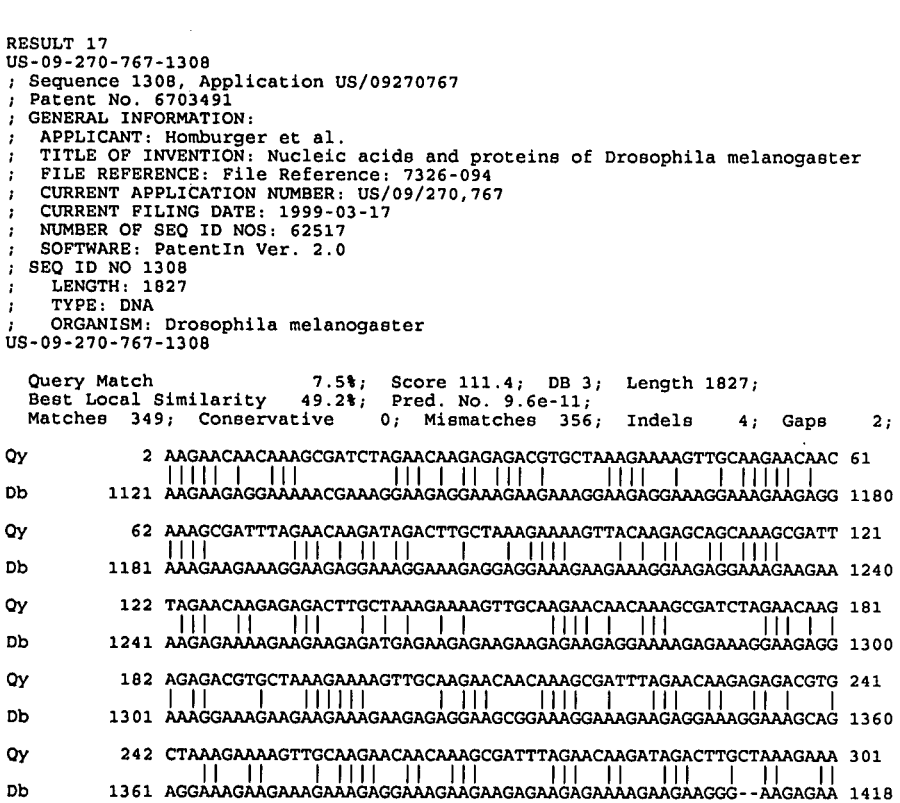
Qy        977 ATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGG 1036
              |   |        || | ||  |  |||||   | |||  | |     |   | | ||| |
Db       2084 AACAAAATATACAAGAAGATGATAATGATGAAGAGGATGATGATGATATAGATTTTGAAG 2143

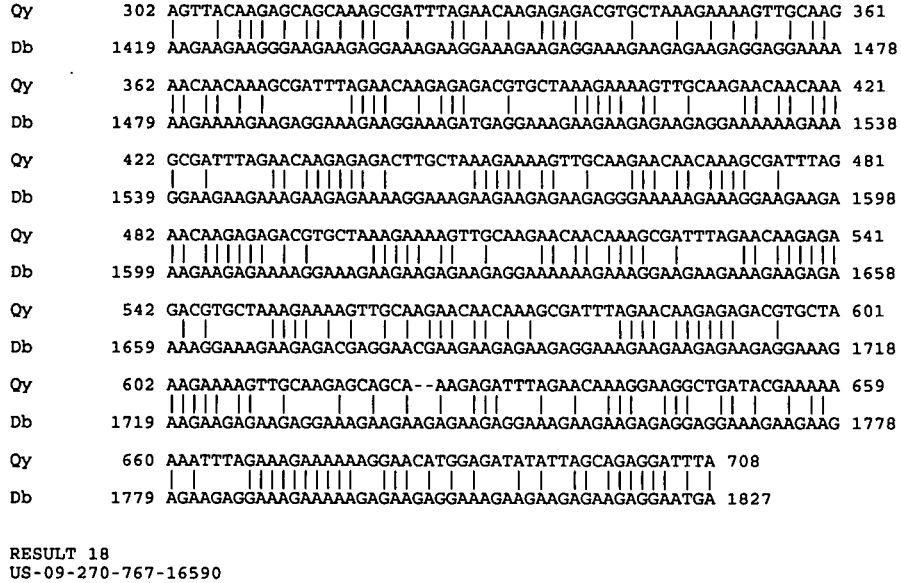
Qy       1037 ATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACT 1096
```



RESULT 17
US-09-270-767-1308
; Sequence 1308, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1308
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1308

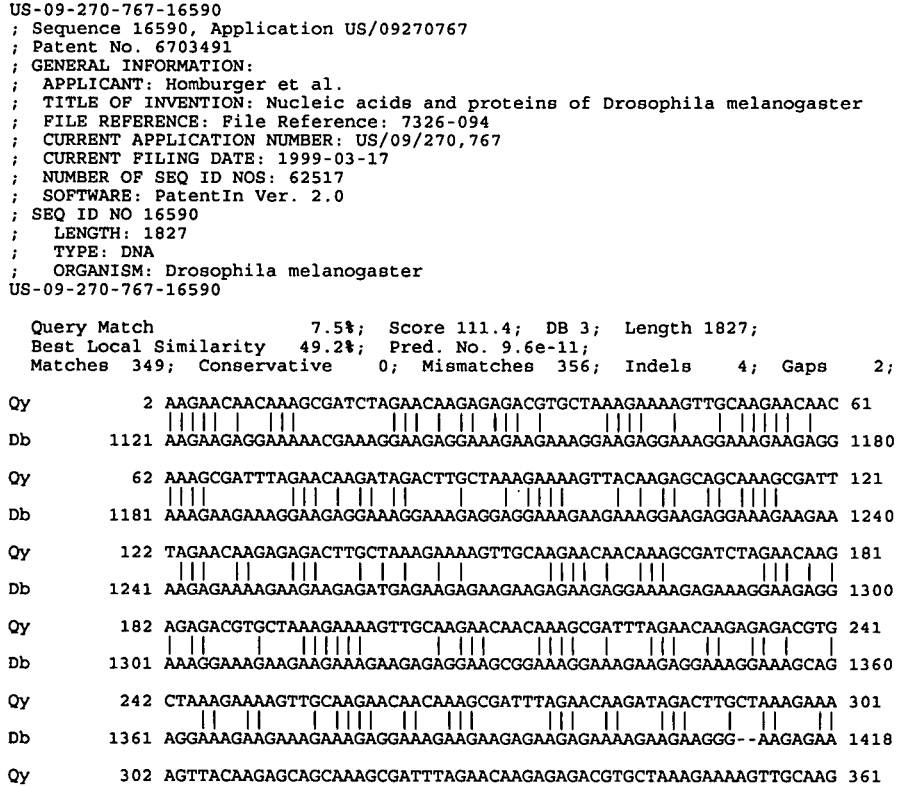
Query Match 7.5%; Score 111.4; DB 3; Length 1827;
Best Local Similarity 49.2%; Pred. No. 9.6e-11;
Matches 349; Conservative 0; Mismatches 356; Indels 4; Gaps 2;





RESULT 18
US-09-270-767-16590
; Sequence 16590, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16590
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16590

Query Match 7.5%; Score 111.4; DB 3; Length 1827;
Best Local Similarity 49.2%; Pred. No. 9.6e-11;
Matches 349; Conservative 0; Mismatches 356; Indels 4; Gaps 2;



```
                |   | |||  | || | ||  |   || | |  ||||    |   |  | | |  | ||
Db       1419 AAGAAGAAGGGAAGAAGAGGAAAGAAGGAAAGAAGAGGAAAGAAGAGAAGAGGAGGAAAA 1478

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
              || || | |        ||||  | |||   |     ||||| ||  |    || || |||
Db       1479 AAGAAAAGAAGAGGAAAGAAGGAAAGATGAGGAAAGAAGAAGAGAAGAGGAAAAAAGAAA 1538

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
              |  |     || ||||||  |       |||||  ||  |   ||| || ||||  |
Db       1539 GGAAGAAGAAAGAAGAGAAAAGGAAAGAAGAAGAGAAGAGGGAAAAAGAAAGGAAGAAGA 1598

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
              || ||||||  | |     |||||  ||  |    || || ||||  |     || ||||||
Db       1599 AAGAAGAGAAAAGGAAAGAAGAAGAGAAGAGGAAAAAAGAAAGGAAGAAGAAAGAAGAGA 1658

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
               | |     |||| | |  | | ||| || | |       |||| ||||||   |
Db       1659 AAAGGAAAGAAGAGACGAGGAACGAAGAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAG 1718

Qy        602 AAGAAAAGTTGCAAGAGCAGCA--AAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAA 659
              ||||| ||  |    |  | |  |  |||   |  |  |||  |||  || |  | ||
Db       1719 AAGAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAAAGAAGAAGAGAGGAGGAAAGAAGAAG 1778

Qy        660 AAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
              | |    ||||||||||  ||| | | |   | |  || ||||||  | |
Db       1779 AGAAGAGGAAAGAAAAAGAGAAGAGGAAAGAAGAAGAGAAGAGGAATGA 1827


RESULT 19
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             7.4%;  Score 110.2;  DB 2;  Length 7218;
  Best Local Similarity   4.1%;  Pred. No. 1.7e-10;
  Matches   16;  Conservative  265;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||    | ||    | |   |||| | ||::::::: : :  :::::::::  : ::
Db       1465 AAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1406

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ::: :: :::: ::   :::: :::::::: : :  :::::::::  : ::::: :: ::
Db       1405 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1346

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              :: ::   :::: :::::::::   :  :::::::::  : ::::: :: :::: ::   :
Db       1345 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1286

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ::: :::::::: : :  :::::::::  : ::::: :: :::: ::   :::: :::::
Db       1285 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1226

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ::: : :  :::::::::  : ::::: :: :::: ::   :::: :::::::: : :
Db       1225 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1166

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              :::::::::  : ::::: :: :::::::   :::: :::::::::  :: : :::::::
Db       1165 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1106

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGA 689
              ::   ::::::::::::::::: : :::::
Db       1105 RRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1077


RESULT 20
US-08-559-896B-1
; Sequence 1, Application US/08559896B
; Patent No. 6310046
;  GENERAL INFORMATION:
;    APPLICANT:  Patrick E. Duffy
;    APPLICANT:  Christian F. Ockenhouse
;    TITLE OF INVENTION:  SEQUESTRIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John Moran
;      STREET:  USA MRMC -  MCMR-JA
;      CITY:  FORT DETRICK, FREDERICK
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21702-5012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,896B
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Moran, John
;      REGISTRATION NUMBER:  26,313
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-7714
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1956 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
US-08-559-896B-1

  Query Match             6.7%;  Score 99.6;  DB 3;  Length 1956;
  Best Local Similarity   43.6%;  Pred. No. 1.1e-08;
  Matches  493;  Conservative    0;  Mismatches  634;  Indels    3;  Gaps    1;

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 406
              |||||  ||    ||||  |   ||       | |||||||| | |  |   ||    ||
Db        293 ATGAAATGTATAGAGAAGAGTTAAATAGGATTGAACAAGAAAAAATTAGAAATATGAATA 352

Qy        407 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAAC 466
              |||||  | | |    ||   | | || | |  |  ||   | |    ||||
Db        353 AAGAAGAAATAAATAAAACTTACAAGGACGAAATAAATAATATGAATAGTGATCAAGTTG 412

Qy        467 AACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 526
              |  |||   ||  ||||  ||      |     || || |      | |  ||
Db        413 ATAAAATACATAGAGAAGAATTAGAAAAAATCGAAAAAGAAAAAATAAACAAAATGGATA 472

Qy        527 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 586
              |    |||  |||  | |      || ||||  | ||  |  ||     |  | |||  |
Db        473 AAGATGAAATAGATAAAATATATAGAGAAGAATTAGACAAAATGGATCGAGATGCAATTT 532

Qy        587 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGG 646
              |    |    |    ||||  |      |  |  |  || |||  |||    |||   |
Db        533 ATAGTATGTATATAGAAGATATAAGTAACAAAAATATAAAAGACTTAATTAAAAATGAAA 592

Qy        647 CTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706
              || || || ||  ||  | | | ||||||  ||     |     |    |  | |
Db        593 AGGAAACAAATAAGGATAAAAACAAAAAAAAAGATATAGATATAAACAAAAAGAAAAAAA 652

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766
              | ||     | ||||  ||          ||  |  |    || ||  |    | | | |
Db        653 AAGATATAGATATAGATGTAGACATAGATAAAGATATACATAAAGATCATGTAGAAGAAT 712

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826
              | | |     | |||  |  |  |  |    |  |||| |  ||     | |  |
Db        713 TATACGGAGAAGTAAAAAACAAACTTAGCAAAGAAGAATTAGATAGAATGGACAGAGACG 772

Qy        827 AAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886
              | | | || | || |   | | |   |||   || | | | | |    |     ||
Db        773 CATTATATAGAGTATACCTAGAAGAATTAGATAGAATGAACAGAGACGAATTATATAGAG 832

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946
              |  || |  ||  | |      ||| ||| ||||      || ||  |    | |||
Db        833 TATACCTAGAAGAATTAGAAAAAATAGACAAGGAAGAAAAAGAAAAAATTCATAGAGAAA 892

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006
              ||  | | |||  ||    |     |||  | | || | ||  |  || |  | |
Db        893 AATTACACAAAATTGAAAAAGAGAAAATAAATAAAATGGATAAAGATCAAATAGATAAAA 952

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066
              |  ||||     || |||  ||   ||  |||   ||    | || |||        |
Db        953 TATATGAAGAAGAATTAAACAAAATGGATAGTGATGAAA---TTCAACATGTAAGGAGAG 1009

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126
              ||||    ||||   | |  | |||  |  || |  | ||  | |      |      ||
Db       1010 CAATATTACAAGATATACAAAAAGAAAAAATACAAAATTTAGAACTAGAAGAAATCGATA 1069

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186
              ||   |||| |||  | | | |    |   |||  |    | ||  |  |    ||
Db       1070 GACTTTATAAAGAAGAATTAGATAGAATGGATAGAGAAGCAAGGTATGAAATCCCCATGA 1129

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246
              ||||  ||   | ||||  |||| |   ||| | | |      || |   ||||
Db       1130 GAAATTTAAGCAGAAATGAAAAAGATAATATTATACATAGAAATATTAAAAATGAATCTA 1189

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306
              |  ||||||  || |||| |||   | ||| |   | ||   |  | |||  | |    |

Db       1190 ACCAAAAAAATAAGAAAGAAAATGTAAATGTATTTATAATACACGACAATAATGATAGTA 1249

Qy       1307 ATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATA 1366
              || |||| | |    || || |      |   ||  ||  |||| |||||||    ||||
Db       1250 ATAATAATAATAATAATAATAATAGAGATGTGAATAATTTAAATAATAAGCACACAAATA 1309

Qy       1367 AGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTT 1426
              |  |  |  |  ||    || ||      ||   | ||    |  ||||| || |
Db       1310 ATAATTATAATGAAAATGTAGAAGTTGAATTAGTTGTACGAAATTTAGACAAGGATAAAG 1369

Qy       1427 TACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAA 1476
              |   |   | || ||  |    ||  || | |  ||| |  |||  |||
Db       1370 GAGCCAAGATAGAAGATATTATAGACTATTTTAACAAAGAAATTAAAAAA 1419


RESULT 21
US-09-351-794A-1
; Sequence 1, Application US/09351794A
; Patent No. 6641815
; GENERAL INFORMATION:
;  APPLICANT: DUFFY, PATRICK E.
;  APPLICANT:  OCKENHOUSE, CHRISTIAN F.
;  TITLE OF INVENTION: SEQUESTRIN
;  FILE REFERENCE: 38644-175519
;  CURRENT APPLICATION NUMBER: US/09/351,794A
;  CURRENT FILING DATE:  1999-07-13
;  PRIOR APPLICATION NUMBER: 08/559,896
;  PRIOR FILING DATE: 1995-11-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1956
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-351-794A-1

  Query Match             6.7%;  Score 99.6;  DB 3;  Length 1956;
  Best Local Similarity   43.6%;  Pred. No. 1.1e-08;
  Matches  493;  Conservative    0;  Mismatches  634;  Indels    3;  Gaps    1;

Qy        347 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 406
              |||||  ||    ||||  |   ||       | |||||||| | |  |   ||    ||
Db        293 ATGAAATGTATAGAGAAGAGTTAAATAGGATTGAACAAGAAAAAATTAGAAATATGAATA 352

Qy        407 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAAC 466
              |||||  | | |    ||   | | || | |  |  ||   | |    ||||
Db        353 AAGAAGAAATAAATAAAACTTACAAGGACGAAATAAATAATATGAATAGTGATCAAGTTG 412

Qy        467 AACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 526
              |  |||   ||  ||||  ||      |     || || |      | |  ||
Db        413 ATAAAATACATAGAGAAGAATTAGAAAAAATCGAAAAAGAAAAAATAAACAAAATGGATA 472

Qy        527 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 586
              |    |||  |||  | |      || ||||  | ||  |  ||     |  | |||  |
Db        473 AAGATGAAATAGATAAAATATATAGAGAAGAATTAGACAAAATGGATCGAGATGCAATTT 532

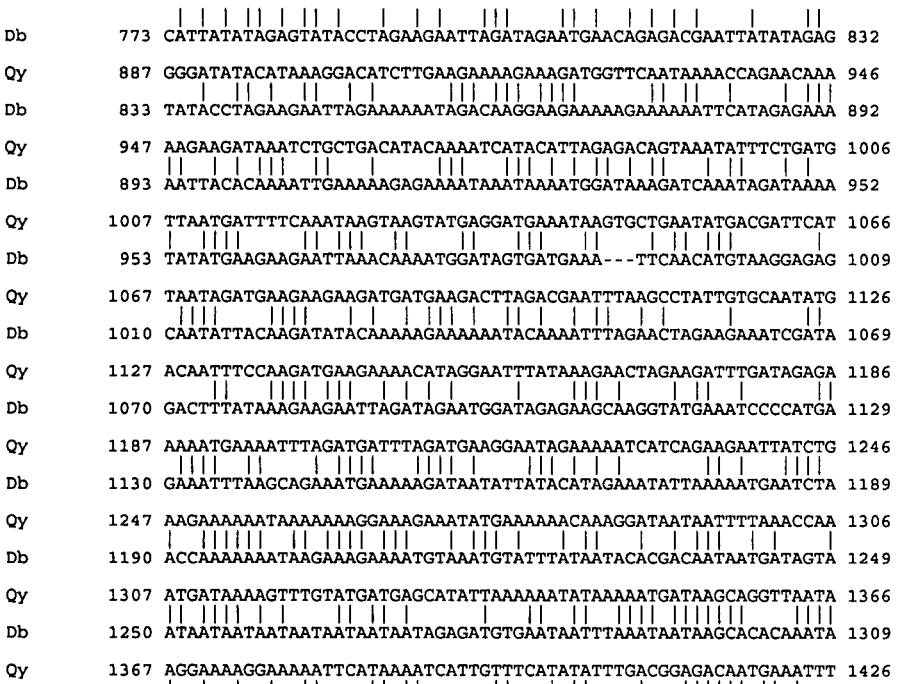
Qy        587 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGG 646
              |    |    |    ||||  |      |  |  |  || |||  |||    |||   |
Db        533 ATAGTATGTATATAGAAGATATAAGTAACAAAAATATAAAAGACTTAATTAAAAATGAAA 592

Qy        647 CTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATT 706
              || || || ||  ||  | | | ||||||  ||     |     |    |  | |
Db        593 AGGAAACAAATAAGGATAAAAACAAAAAAAAAGATATAGATATAAACAAAAAGAAAAAAA 652

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766
              | ||     | ||||  ||          ||  |  |    || ||  |    | | | |
Db        653 AAGATATAGATATAGATGTAGACATAGATAAAGATATACATAAAGATCATGTAGAAGAAT 712

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826
              | | |     | |||  |  |  |  |    |  |||| |  ||     | |  |
Db        713 TATACGGAGAAGTAAAAAACAAACTTAGCAAAGAAGAATTAGATAGAATGGACAGAGACG 772

Qy        827 AAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886
```



RESULT 22
US-09-902-540-1318
; Sequence 1318, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1318
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(614)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1318

Query Match 6.3%; Score 93.6; DB 3; Length 614;
Best Local Similarity 47.7%; Pred. No. 1e-07;
Matches 270; Conservative 0; Mismatches 296; Indels 0; Gaps 0;



RESULT 23
US-09-949-016-12776
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match 6.3%; Score 93.4; DB 3; Length 187169;
Best Local Similarity 46.8%; Pred. No. 1.9e-07;

```
  Matches  363;  Conservative    0;  Mismatches  406;  Indels    6;  Gaps    2;

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766
              |||||         | | |||  |  |||| ||  |  |  | | |    |      ||||
Db        466 TATATATATAAAAATATATAATATATATAAAATATATAAAAAATATATATAAAATATATA 525

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826
                |||     |   || |    ||  | |      | | |    ||| || | | |  |
Db        526 TAATATATAAAATATATATAATATATAATATATATAAAATATATATTATATAGTATATAT 585

Qy        827 AAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886
              |||| | |||| || | || | | ||| | |          | ||    | ||  |   |
Db        586 AAATATATATACTATATAATATATATAAATATATATAATATATAATATATAAAATATATA 645

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946
                  ||| |  | || |  | | |   || ||   | ||  | |    || | |
Db        646 TAATATATAATATATAAAATATATATAATATATAATATACATAAAATATATATAAATATA 705

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006
                | | ||| |       | |    ||| || | | | |  |||||| |    ||
Db        706 TAATATATATAAAAATATAAAATATATAATATATATAAAAAATATAAAATATATAATA 765

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066
              |  |  || |  ||   |   |||    || | |||     |  |||  |    | |
Db        766 TATAAAATATAAAATATATATTATAT----ATAATATATAAAATATATAAAATATATAAA 821

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126
                |||  |   |  || ||    || |  || |  |  | ||   |||      |
Db        822 ATATATAATATATAAAATATATATAATATATAAAATATATAAATATATATAAAATATAT 881

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186
              | ||| |  || ||  | || | |||  || |    || |  || || | ||| | | |
Db        882 AAAATATATAAAATATATAATATATATAAAATATATAATATATATAATATATAATATATA 941

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246
              |||   | ||| || |  | || ||| |   |||  | |      | |  || ||| |
Db        942 AAATATATAATATATAAAAAATATATGGATATATAATATATATGATATATAAAATATATA 1001

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306
              |   | | | || | ||   | ||||||  || |   ||     |||||  |  ||   |
Db       1002 ATATATATAAAATATAAAATATAAAATATATAATATATAAAATATATAATATATAATATA 1061

Qy       1307 ATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATA 1366
                | |   | ||| ||    ||| || | |||||  ||   ||||      |||||
Db       1062 TAATGTATAAAATATATAATAT--ATAATATATAATATATAATATATAAAATATATAATA 1119

Qy       1367 AGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTT 1426
                || |   | |||  |||| ||  |   |  ||| || |     | |||   | |||
Db       1120 TATAATATATATAATATATAATATATTATATATATAAATATATATAATACATATATATTA 1179

Qy       1427 TACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATA 1481
              || | ||  |  || |  |||    | |||||   | ||| | |||  |  | ||
Db       1180 TATATATTATATATAATATATAAATATATATATAAATATAATATATATATATTTA 1234


RESULT 24
US-09-949-016-15940
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
;   LENGTH: 191569
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(191569)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

  Query Match             6.3%;  Score 93.4;  DB 3;  Length 191569;
  Best Local Similarity   46.8%;  Pred. No. 1.9e-07;
  Matches  363;  Conservative    0;  Mismatches  406;  Indels    6;  Gaps    2;

Qy        707 TATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATT 766
              |||||         | | |||  |  |||| ||  |  |  | | |    |      ||||
Db        466 TATATATATAAAAATATATAATATATATAAAATATATAAAAAATATATATAAAATATATA 525

Qy        767 ATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGG 826
                |||     |   || |    ||  | |      | | |    ||| || | | |  |
Db        526 TAATATATAAAATATATATAATATATAATATATATAAAATATATATTATATAGTATATAT 585

Qy        827 AAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAA 886
              |||| | |||| || | || | | ||| | |          | ||    | ||  |   |
Db        586 AAATATATATACTATATAATATATATAAATATATATAATATATAATATATAAAATATATA 645

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946
                  ||| |  | || |  | | |   || ||   | ||  | |    || | |
Db        646 TAATATATAATATATAAAATATATATAATATATAATATACATAAAATATATATAAATATA 705

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006
                | | ||| |       | |    ||| || | | | |  |||||| |    ||
Db        706 TAATATATATAAAAATATAAAATATATAATATATATAAAAAATATAAAATATATAATA 765

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066
              |  |  || |  ||   |   |||    || | |||     |  |||  |    | |
Db        766 TATAAAATATAAAATATATATTATAT----ATAATATATAAAATATATAAAATATATAAA 821

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126
                |||  |   |  || ||    || |  || |  |  | ||   |||      |
Db        822 ATATATAATATATAAAATATATATAATATATAAAATATATAAATATATATAAAATATAT 881

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGA 1186
              | ||| |  || ||  | || | |||  || |    || |  || || | ||| | | |
Db        882 AAAATATATAAAATATATAATATATATAAAATATATAATATATATAATATATAATATATA 941

Qy       1187 AAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTG 1246
              |||   | ||| || |  | || ||| |   |||  | |      | |  || ||| |
Db        942 AAATATATAATATATAAAAAATATATGGATATATAATATATATGATATATAAAATATATA 1001

Qy       1247 AAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAA 1306
              |   | | | || | ||   | ||||||  || |   ||     |||||  |  ||   |
Db       1002 ATATATATAAAATATAAAATATAAAATATATAATATATAAAATATATAATATATAATATA 1061

Qy       1307 ATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATA 1366
                | |   | ||| ||    ||| || | |||||  ||   ||||      |||||
Db       1062 TAATGTATAAAATATATAATAT--ATAATATATAATATATAATATATAAAATATATAATA 1119

Qy       1367 AGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTT 1426
                || |   | |||  |||| ||  |   |  ||| || |     | |||   | |||
Db       1120 TATAATATATATAATATATAATATATTATATATATAAATATATATAATACATATATATTA 1179

Qy       1427 TACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATA 1481
              || | ||  |  || |  |||    | |||||   | ||| | |||  |  | ||
Db       1180 TATATATTATATATAATATATAAATATATATATAAATATAATATATATATATTTA 1234


RESULT 25
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
```

```
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 5361
;   TYPE: DNA
;   ORGANISM: P. falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5361)
US-08-973-462-2

  Query Match             6.2%;  Score 92.6;  DB 3;  Length 5361;
  Best Local Similarity   45.5%;  Pred. No. 1.9e-07;
  Matches  583;  Conservative    0;  Mismatches  669;  Indels   29;  Gaps    6;

Qy         90 TAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAA 149
              ||  ||| ||      |  | || |    | |  |  | | || |   |  |||   |
Db       3672 TATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTAT 3731

Qy        150 GTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 209
             ||   ||| |     ||  || | |||          |||| | |    ||
Db       3732 TGAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATT 3791

Qy        210 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 269
              | || ||   ||   || | |     | |||||||||   |  | | | || |
Db       3792 AAAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGA 3851

Qy        270 CGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGA 329
              |||||| ||| ||     |  || |  |  ||     |  |      ||  | |
Db       3852 TACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGT 3911

Qy        330 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 389
              | ||   | |  ||  ||||| |   |   ||| ||  |  |||||| || |
Db       3912 AGAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGA------ 3965

Qy        390 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAA 449
              | ||   |||| |  ||  |||||  | ||        |||||  || ||       |
Db       3966 TCTTGAAGAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTT 4025

Qy        450 AGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 509
              |||| | |   |||||  | |||     |  |  | |       |   |  ||| || | |
Db       4026 AGAAGATTATAAAGAATTAAAAACTATTGAAACAGATATTTTAGAAGAGAAAAAAGAAAT 4085

Qy        510 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 569
              || || | |      |||| ||| ||      |   |     ||| |  |  |||| |
Db       4086 AGAAAAAGATC------ATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCT 4139

Qy        570 ACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGA 629
              ||   ||| ||  | ||||   |  | |   ||||       ||||  |  ||  |||
Db       4140 TGAAGCAGATATATTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGA 4199

Qy        630 TTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGA 689
              || |||          ||   |  | || || ||  |||      | |  |||  | |
Db       4200 AGAAGTACACGAATTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAA 4259

Qy        690 TATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGA 749
              || | ||| |||   |  |      ||||      |   | || |  |  || |||
Db       4260 AGGTTTGGAAGAAGATGATTTAGAAGAAGTAGATGATTTA---AAAGGAAGTATATTAGA 4316

Qy        750 AAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAA 809
              |    |  ||| | |||  |         || |   | |    ||   |  ||| | ||
Db       4317 CATGTTAAAGGGAGA-TATGGAATTAGGGGATATGGATAAGGAAAGTTTAGAAGATGTAA 4375

Qy        810 TAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTA-CAA 868
              |  | |  ||  |   | |  |  ||  |  ||| |      ||   | |  |||   |
Db       4376 CAACAAAACTTGGAGAAAGAGTTGAATCCTTAAAAGATGTTTTATCTAGTGCATTAGGCA 4435

Qy        869 CAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTT 928
                |||  |||  |   |      | | | ||||  ||      |   | ||||  |  |
Db       4436 TGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTGGAAGAAGTAT 4495

Qy        929 CAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGA 988
              | |||||  |||     ||||||||   |       | |||  |||  | | | ||
Db       4496 TATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAGAAAGTA---- 4551

Qy        989 CAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTG 1048
                  | | |||| |||| |||     ||  ||   |  |    |  ||| |
Db       4552 --------AGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATAGTAGAAGTTGAAA 4603

Qy       1049 CTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTA 1108
               || |||| |||  |  |  ||||| ||||||||||   |||||        |||
Db       4604 TGAAAGATGAAGATATAGAAGAAGATGTAGAAGAAGATATAGAAGAAGATATAGAAGAAG 4663

Qy       1109 AGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAAC 1168
              |      |||   |     ||  |       ||||||||| |      | |    |  | |||
Db       4664 ATAAAGTTGAAGATATAGATGAAGATATAGATGAAGATATAGGTGAAGACAAAGATGAAG 4723

Qy       1169 TAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAAT 1228
              |   |||||| ||||   |||| || ||       |||    | | |||  | | | |||
Db       4724 TTATAGATTTAATAGTCCAAAAAGAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAA 4783

Qy       1229 CATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGG 1288
              || |||| ||  |  |||||||    | |          || ||| | | | |
Db       4784 AATTAGAAAAAAAAGTTGAAGAAGGTGTTAGTGGTCTTAAAAAACACGTAGACGAAGTAA 4843

Qy       1289 ATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348
              ||  || || ||  || |||||||     |  || | ||      ||||   |  ||
Db       4844 TGAAATATGTTCAAAAAATTGATAAAGAAGTTGATAAAGAAGTATCTAAAGCTTTAGAAT 4903

Qy       1349 ATGATAAGCAGGTTAATAAGG 1369
                | ||  | |||| ||| |
Db       4904 CAAAAAATGATGTTACTAATG 4924


RESULT 26
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US-08-973-462-1

  Query Match             6.2%;  Score 92.6;  DB 3;  Length 6152;
  Best Local Similarity   45.5%;  Pred. No. 1.9e-07;
  Matches  583;  Conservative    0;  Mismatches  669;  Indels   29;  Gaps    6;

Qy         90 TAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAA 149
              ||  ||| ||      |  | || |    | |  |  | | || |   |  |||   |
Db       3916 TATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTAT 3975

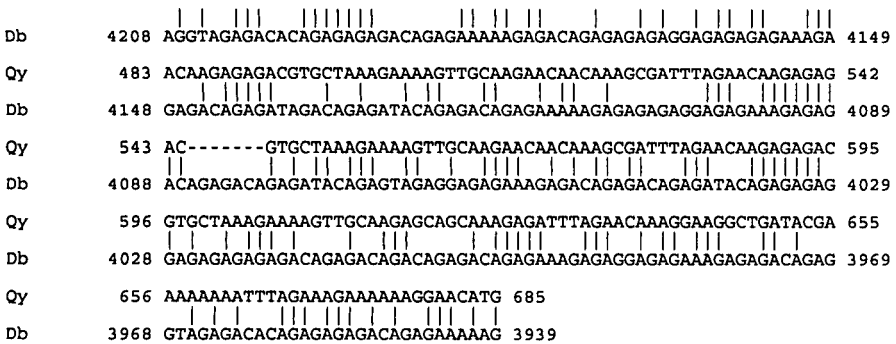
Qy        150 GTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 209
```

```
Db       3976 TGAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATT 4035
Qy        210 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 269
Db       4036 AAAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGA 4095
Qy        270 CGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGA 329
Db       4096 TACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGT 4155
Qy        330 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 389
Db       4156 AGAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGA------ 4209
Qy        390 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAA 449
Db       4210 TCTTGAAGAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTT 4269
Qy        450 AGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTT 509
Db       4270 AGAAGATTATAAAGAATTAAAAACTATTGAAACAGATATTTTAGAAGAGAAAAAAGAAAT 4329
Qy        510 GCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACA 569
Db       4330 AGAAAAAGATC------ATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCT 4383
Qy        570 ACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGA 629
Db       4384 TGAAGCAGATATATTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGA 4443
Qy        630 TTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGA 689
Db       4444 AGAAGTACACGAATTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAA 4503
Qy        690 TATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGA 749
Db       4504 AGGTTTGGAAGAAGATGATTTAGAAGAAGTAGATGATTTA---AAAGGAAGTATATTAGA 4560
Qy        750 AAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAA 809
Db       4561 CATGTTAAAGGGAGA-TATGGAATTAGGGGATATGGATAAGGAAAGTTTAGAAGATGTAA 4619
Qy        810 TAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTA-CAA 868
Db       4620 CAACAAAACTTGGAGAAAGAGTTGAATCCTTAAAAGATGTTTTATCTAGTGCATTAGGCA 4679
Qy        869 CAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTT 928
Db       4680 TGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTGGAAGAAGTAT 4739
Qy        929 CAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGA 988
Db       4740 TATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAGAAAGTA---- 4795
Qy        989 CAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTG 1048
Db       4796 --------AGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATAGTAGAAGTTGAAA 4847
Qy       1049 CTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTA 1108
Db       4848 TGAAAGATGAAGATATAGAAGAAGATGTAGAAGAAGATATAGAAGAAGATATAGAAGAAG 4907
Qy       1109 AGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAAC 1168
Db       4908 ATAAAGTTGAAGATATAGATGAAGATATAGATGAAGATATAGGTGAAGACAAAGATGAAG 4967
Qy       1169 TAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAAT 1228
Db       4968 TTATAGATTTAATAGTCCAAAAAGAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAA 5027
Qy       1229 CATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGG 1288
Db       5028 AATTAGAAAAAAAAAGTTGAAGAAGGTGTTAGTGGTCTTAAAAAACACGTAGACGAAGTAA 5087
Qy       1289 ATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348
Db       5088 TGAAATATGTTCAAAAAATTGATAAAGAAGTTGATAAAGAAGTATCTAAAGCTTTAGAAT 5147
Qy       1349 ATGATAAGCAGGTTAATAAGG 1369
Db       5148 CAAAAAATGATGTTACTAATG 5168
```

RESULT 27
US-09-949-016-11814/c
; Sequence 11814, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11814
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11814

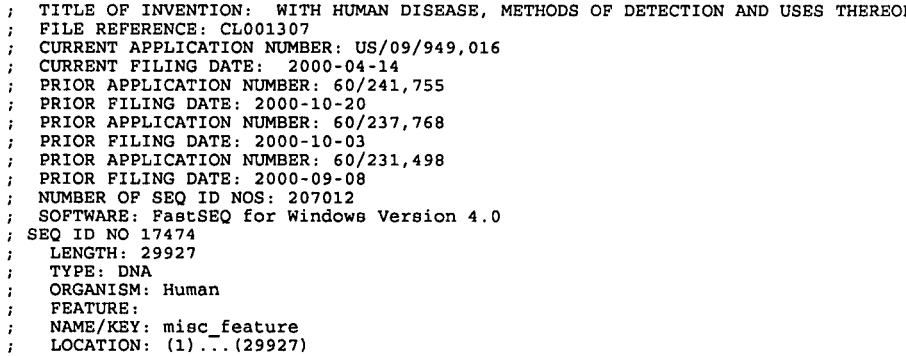
Query Match 6.2%; Score 91.6; DB 3; Length 29927;
Best Local Similarity 47.0%; Pred. No. 3.2e-07;
Matches 324; Conservative 0; Mismatches 359; Indels 7; Gaps 1;

```
Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
Db       4628 AGACACAGAGAGAGACAGAGAAAAAGAGACAGAGAGGAGAGAGACAAAGAGAGACAGAGA 4569
Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
Db       4568 CAGACAGAGAGACAGAGACAGAAAAAGAGAGAGAGGAGAGAGAGAAAGAGAGACAGAGAT 4509
Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
Db       4508 AGAGATACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAGAGGTAGAGACACAGA 4449
Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
Db       4448 GAGAGACAGAGAAAAAGAGACAGAGAGAGATGAGAGAGAGAAAGAGAGACAGAGATAGAC 4389
Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
Db       4388 AGAGATACAGAGACAGAGAAAAAGAGAGAGAGGAGAGAAAGAGAGACAGAGTCAGAGATA 4329
Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
Db       4328 CAGAGAGAGGAGAGAAAGAGAGACAGAGACAGAGATACAGAGAGAGGAGAGAAAGAGAGA 4269
Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
Db       4268 CAGAGACAGACAGAGATACAGACAGAGACAGAGAAAAAGAGAGGAGAGAAAGAGAGACAG 4209
Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
```



RESULT 28
US-09-949-016-17474/c
; Sequence 17474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17474
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17474

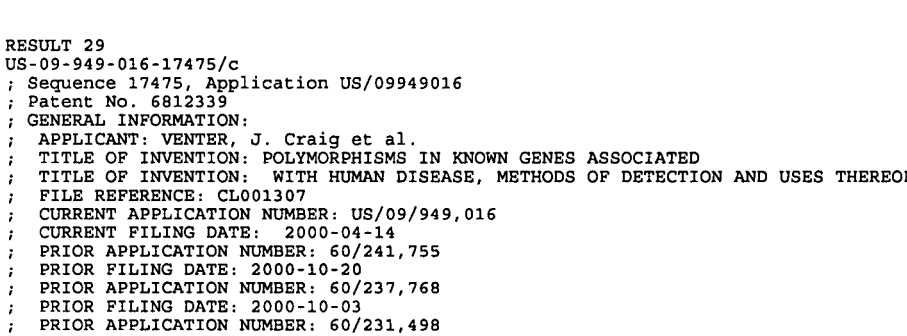
Query Match 6.2%; Score 91.6; DB 3; Length 29927;
Best Local Similarity 47.0%; Pred. No. 3.2e-07;
Matches 324; Conservative 0; Mismatches 359; Indels 7; Gaps 1;

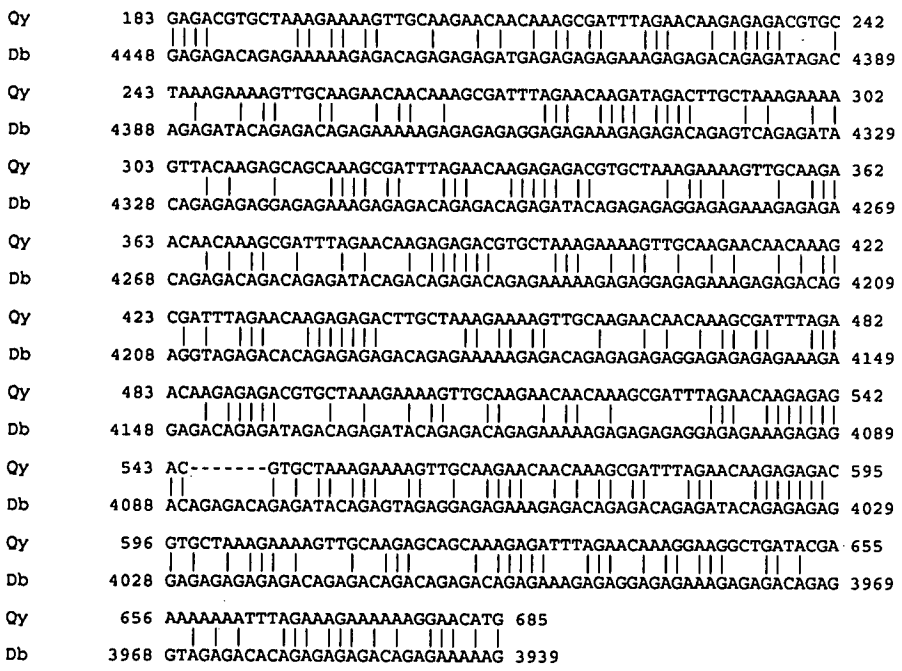




RESULT 29
US-09-949-016-17475/c
; Sequence 17475, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17475
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17475

Query Match 6.2%; Score 91.6; DB 3; Length 29927;
Best Local Similarity 47.0%; Pred. No. 3.2e-07;
Matches 324; Conservative 0; Mismatches 359; Indels 7; Gaps 1;

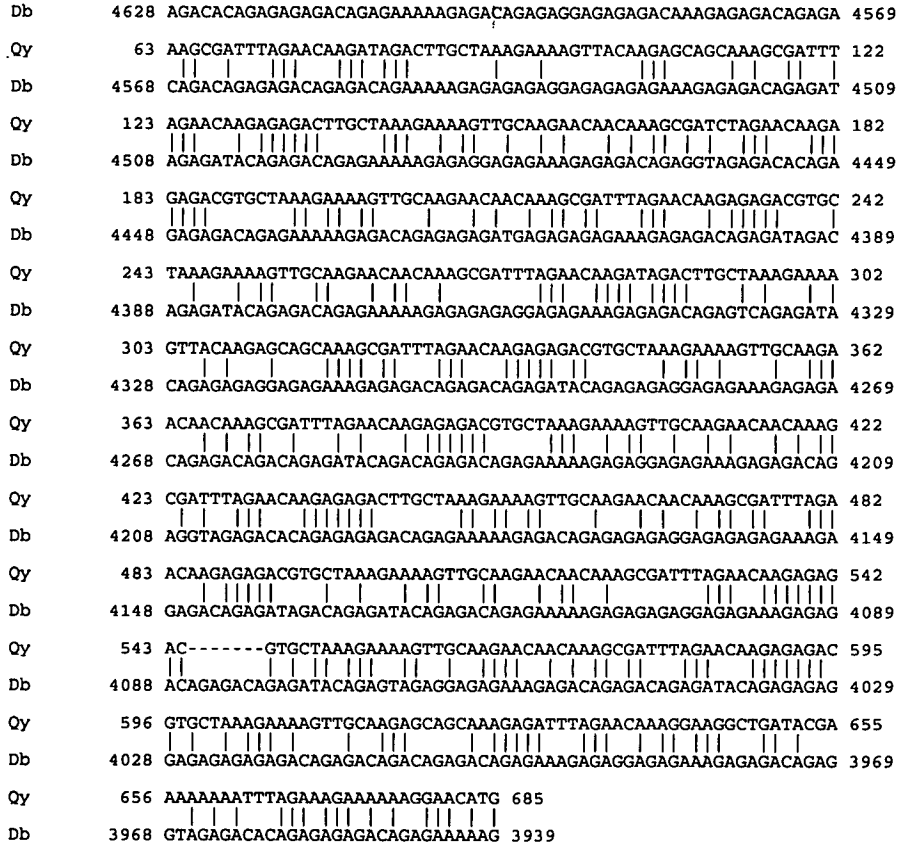




RESULT 30
US-09-949-016-12639/c
; Sequence 12639, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12639
; LENGTH: 37802
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(37802)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12639

Query Match 6.2%; Score 91.6; DB 3; Length 37802;
Best Local Similarity 47.0%; Pred. No. 3.3e-07;
Matches 324; Conservative 0; Mismatches 359; Indels 7; Gaps 1;





Search completed: December 20, 2005, 15:19:21
Job time : 344.317 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 10:43:25 ; Search time 1605.5 Seconds
(without alignments)
7633.272 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	1482	100.0	1482	3	US-09-837-344-45	Sequence 45, Appl
	2	1482	100.0	1482	3	US-09-900-963-45	Sequence 45, Appl
	3	1480.4	99.9	1482	3	US-09-837-344-41	Sequence 41, Appl
	4	1480.4	99.9	1482	3	US-09-900-963-41	Sequence 41, Appl
	5	1444.4	97.5	1493	3	US-09-837-344-38	Sequence 38, Appl
	6	1444.4	97.5	1493	3	US-09-900-963-38	Sequence 38, Appl
	7	948.4	64.0	950	3	US-09-837-344-32	Sequence 32, Appl
	8	948.4	64.0	950	3	US-09-900-963-32	Sequence 32, Appl
	9	472.8	31.9	988	3	US-09-837-344-34	Sequence 34, Appl
	10	472.8	31.9	988	3	US-09-900-963-34	Sequence 34, Appl
	11	470.8	31.8	954	3	US-09-837-344-37	Sequence 37, Appl
	12	470.8	31.8	954	3	US-09-900-963-37	Sequence 37, Appl
	13	231.8	15.6	1300	7	US-10-668-749A-1	Sequence 1, Appli
c	14	151.6	10.2	1297	8	US-10-425-115-67687	Sequence 67687, A
c	15	151.2	10.2	6292	7	US-10-221-714A-461	Sequence 461, App
c	16	149.2	10.1	6767	7	US-10-221-714A-330	Sequence 330, App
c	17	145	9.8	1204	7	US-10-437-963-77858	Sequence 77858, A
c	18	144.4	9.7	3662	10	US-11-097-143-5152	Sequence 5152, Ap
c	19	144.2	9.7	1243	8	US-10-425-115-172717	Sequence 172717,
c	20	140.8	9.5	1081	8	US-10-425-115-16756	Sequence 16756, A
c	21	139.4	9.4	9539	5	US-10-239-676-52	Sequence 52, Appl
c	22	139.4	9.4	9539	6	US-10-240-453-54	Sequence 54, Appl
c	23	136.2	9.2	6668	6	US-10-311-455-1670	Sequence 1670, Ap
c	24	131.2	8.9	778	8	US-10-363-345A-2179	Sequence 2179, Ap
	25	131.2	8.9	778	8	US-10-363-345A-2180	Sequence 2180, Ap
c	26	131.2	8.9	778	9	US-10-363-483A-2179	Sequence 2179, Ap
	27	131.2	8.9	778	9	US-10-363-483A-2180	Sequence 2180, Ap
c	28	129.8	8.8	7442	7	US-10-221-714A-409	Sequence 409, App
c	29	129	8.7	7597	6	US-10-311-455-986	Sequence 986, App
c	30	128.2	8.7	921	8	US-10-425-115-38710	Sequence 38710, A
c	31	128.2	8.7	1762	8	US-10-363-345A-7529	Sequence 7529, Ap
	32	128.2	8.7	1762	8	US-10-363-345A-7530	Sequence 7530, Ap
c	33	128.2	8.7	1762	9	US-10-363-483A-7529	Sequence 7529, Ap
	34	128.2	8.7	1762	9	US-10-363-483A-7530	Sequence 7530, Ap
c	35	127.6	8.6	1995	8	US-10-363-345A-7533	Sequence 7533, Ap
	36	127.6	8.6	1995	8	US-10-363-345A-7534	Sequence 7534, Ap
c	37	127.6	8.6	1995	9	US-10-363-483A-7533	Sequence 7533, Ap
	38	127.6	8.6	1995	9	US-10-363-483A-7534	Sequence 7534, Ap
c	39	127.4	8.6	1200	8	US-10-363-345A-25801	Sequence 25801, A
	40	127.4	8.6	1200	8	US-10-363-345A-25802	Sequence 25802, A
c	41	127.4	8.6	1200	9	US-10-363-483A-25801	Sequence 25801, A
	42	127.4	8.6	1200	9	US-10-363-483A-25802	Sequence 25802, A
c	43	127.4	8.6	14006	6	US-10-311-455-1931	Sequence 1931, Ap
c	44	127	8.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	45	127	8.6	3673778	6	US-10-312-841-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-837-344-45
; Sequence 45, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 45:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1482 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
```

```
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..1482
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-837-344-45

  Query Match            100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity  100.0%;  Pred. No. 9e-200;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 2
US-09-900-963-45
; Sequence 45, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
```

```
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-045
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 45:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1482 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..1482
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-900-963-45

  Query Match             100.0%;  Score 1482;  DB 3;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 9e-200;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 3
US-09-837-344-41
; Sequence 41, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
```

```
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/837,344
;              FILING DATE: 19-Apr-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/462,625
;              FILING DATE: <Unknown>
;              APPLICATION NUMBER: FR 91 01286
;              FILING DATE: 05-FEB-1991
;         ATTORNEY/AGENT INFORMATION:
;              NAME: McGowan, Malcolm K.
;              REGISTRATION NUMBER: 39,300
;              REFERENCE/DOCKET NUMBER: 010830-078
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 41:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1482 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..1482
;         PUBLICATION INFORMATION:
;              DOCUMENT NUMBER:  WO 92/13884
;              PUBLICATION DATE:  20-AUG-1992
;         SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-837-344-41

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-199;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
              |||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 4
US-09-900-963-41
; Sequence 41, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
```

```
;                 DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 41:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1482 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1482
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-900-963-41

  Query Match             99.9%;  Score 1480.4;  DB 3;  Length 1482;
  Best Local Similarity   99.9%;  Pred. No. 1.5e-199;
  Matches 1481;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 840

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
```

```
Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             ||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482


RESULT 5
US-09-837-344-38
; Sequence 38, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1493 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-837-344-38

  Query Match             97.5%;  Score 1444.4;  DB 3;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-194;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257
```

```
Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479
```

```
RESULT 6
US-09-900-963-38
; Sequence 38, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;  INFORMATION FOR SEQ ID NO: 38:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1493 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-900-963-38

  Query Match             97.5%;  Score 1444.4;  DB 3;  Length 1493;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-194;
  Matches 1478;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 TTAGAACAAGAGAGACTTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 179

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 239

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 299

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 359

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        360 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT-AAGAAAAGTTGCAAGAACAACAA 418

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 478

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 538

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 AGACGTGCT-AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 597

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 657

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 717

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        718 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 777

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        778 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 837

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        838 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 897

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 957

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1017

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1077

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1137
```

```
Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1197

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1257

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1317

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1377

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1437

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1482
             |||||||||||||||||||||||||||||||||||||||||||
Db       1438 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATAA 1479


RESULT 7
US-09-837-344-32
; Sequence 32, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-837-344-32

  Query Match             64.0%;  Score 948.4;  DB 3;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 1.4e-124;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy         71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy        131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy        191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy        251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy        311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy        371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy        431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy        491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy        551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy        611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy        671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy        731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy        791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy        851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy        911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950
```

```
RESULT 8
US-09-900-963-32
; Sequence 32, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 950 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-09-900-963-32

  Query Match             64.0%;  Score 948.4;  DB 3;  Length 950;
  Best Local Similarity   99.9%;  Pred. No. 1.4e-124;
  Matches  949;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          11 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 70
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 AAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 60

Qy          71 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 130
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAG 120

Qy         131 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 190
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTG 180

Qy         191 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 240

Qy         251 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 310
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAG 300

Qy         311 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 370
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 360

Qy         371 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 430
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 420

Qy         431 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 490
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 480

Qy         491 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 550
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 540

Qy         551 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 610
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 600

Qy         611 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 670
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 TGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAA 660

Qy         671 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 730
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAG 720

Qy         731 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 790
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTAC 780

Qy         791 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAA 850
               |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         781 CTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATAGAAAAAACAA 840

Qy         851 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 910
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTG 900

Qy         911 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 950


RESULT 9
US-09-837-344-34
; Sequence 34, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
```

```
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 34:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 988 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-837-344-34

  Query Match             31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity   95.7%;  Pred. No. 1.5e-57;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||||||| |||||||||||||||||| |||| || |
Db         661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
               |||||||||||||||||||| ||||| || |||||||||||||||||| |||| ||
Db         841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy         497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy         557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
               ||||||||||||||||||||||||||||
Db         961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 10
US-09-900-963-34
; Sequence 34, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 34:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 988 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-900-963-34

  Query Match             31.9%;  Score 472.8;  DB 3;  Length 988;
  Best Local Similarity   95.7%;  Pred. No. 1.5e-57;
  Matches  486;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         481 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 540

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               |||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         541 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 600

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 660

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||||||| |||||||||||||||||| |||| || |
Db         661 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 720

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 780

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 840

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
```

```
                ||||||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db        841 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 900

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 960

Qy        557 AGTTGCAAGAACAACAAAGCGATTTAGA 584
                ||||||||||||||||||||||||||||
Db        961 GGTTGCAAGAACAACAAAGCGATTTAGA 988


RESULT 11
US-09-837-344-37
; Sequence 37, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 37:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 954 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..954
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-837-344-37

  Query Match           31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 2.8e-57;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
                |||| | |   |||||||||||||||||| | ||| |||||||||| ||||||||||||||
```

```
Db        449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy        137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
                |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db        509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy        197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy        257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
                |||||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db        629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy        317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy        377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy        437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
                |||||||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db        809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy        497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

Qy        557 AGTTGCAAGAACAACAAAGCGATTTA 582
                ||||||||||||||||||||||||||
Db        929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 12
US-09-900-963-37
; Sequence 37, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 37:
```

```
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 954 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..954
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-900-963-37

  Query Match             31.8%;  Score 470.8;  DB 3;  Length 954;
  Best Local Similarity   95.7%;  Pred. No. 2.8e-57;
  Matches  484;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          77 AAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGAC 136
               |||| | |   |||||||||||||||| | ||| |||||||||| ||||||||||||||
Db         449 AAGAAAAAGCAGCTAAAGAAAAGTTACAGGGGCAACAAAGCGATTCAGAACAAGAGAGAC 508

Qy         137 TTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAG 196
               ||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db         509 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAG 568

Qy         197 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         569 AAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGC 628

Qy         257 AAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGC 316
               |||||||||||||||||||||||||||| |||||||||||||||||||| ||||| || |
Db         629 AAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAAC 688

Qy         317 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 376
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         689 AAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATT 748

Qy         377 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 436
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         749 TAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAG 808

Qy         437 AGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 496
               |||||||||||||||||||| ||||| || |||||||||||||||||||| |||| ||
Db         809 AGAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGATAGACTTG 868

Qy         497 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 556
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         869 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAA 928

Qy         557 AGTTGCAAGAACAACAAAGCGATTTA 582
               |||||||||||||||||||||||||
Db         929 GGTTGCAAGAACAACAAAGCGATTTA 954


RESULT 13
US-10-668-749A-1
; Sequence 1, Application US/10668749A
; Publication No. US20040110205A1
; GENERAL INFORMATION:
;  APPLICANT: Agilent Technologies
;  TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis
;  FILE REFERENCE: 50112-1580
;  CURRENT APPLICATION NUMBER: US/10/668,749A
;  CURRENT FILING DATE:  2003-09-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic construct
US-10-668-749A-1

  Query Match             15.6%;  Score 231.8;  DB 7;  Length 1300;
  Best Local Similarity   48.7%;  Pred. No. 1.5e-23;
  Matches  632;  Conservative    0;  Mismatches  667;  Indels    0;  Gaps    0;

Qy          91 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 150
               ||| ||||   | || |  | |||   |  | || || | | |     ||| ||||
Db           1 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy         151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
               || || || |||   |  | || || | | |     ||| ||||    || ||
Db          61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy         211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
               || |||   |  | || || | | |     ||| ||||    || || || |||
Db         121 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 180

Qy         271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330
               | | || || | | |     ||| ||||  | || |  | |||  |  | ||
Db         181 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240

Qy         331 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
               || | | |     ||| ||||    || || || |||   |  | || || | | |
Db         241 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 300

Qy         391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAA 450
               |||  ||||    || || || |||   |  | || || | | |     |||
Db         301 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 360

Qy         451 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
               ||||    || || || |||   |  | || || | | |     ||| ||||
Db         361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 420

Qy         511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
               || || || |||   |  | || || | | |     ||| ||||    || || ||
Db         421 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 480

Qy         571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
               |||  |  | || || | | |     ||| ||||    || |  | |  ||| | ||
Db         481 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 540

Qy         631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
               | || |||  ||    | |  |||||||   | ||| ||||||  || |  | |
Db         541 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 600

Qy         691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750
               | |  | | |  |  | |         | ||| |  |  | | ||     |  | ||
Db         601 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 660

Qy         751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810
               ||  ||      | | | | | |  ||        |    | | || | |   ||
Db         661 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 720

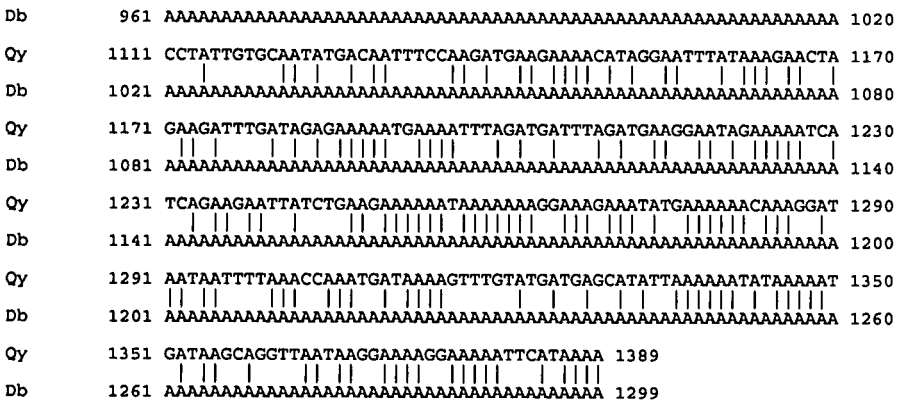
Qy         811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870
               | | | |  ||  |||    | || | ||||| ||| | | ||   |  | || |
Db         721 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 780

Qy         871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930
               ||      ||  |  ||   | | | | |||  | |      || |||| ||| |      |
Db         781 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840

Qy         931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990
               | ||||  | || ||||| || | |||       | | | |||| | | |  | | | |
Db         841 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 900

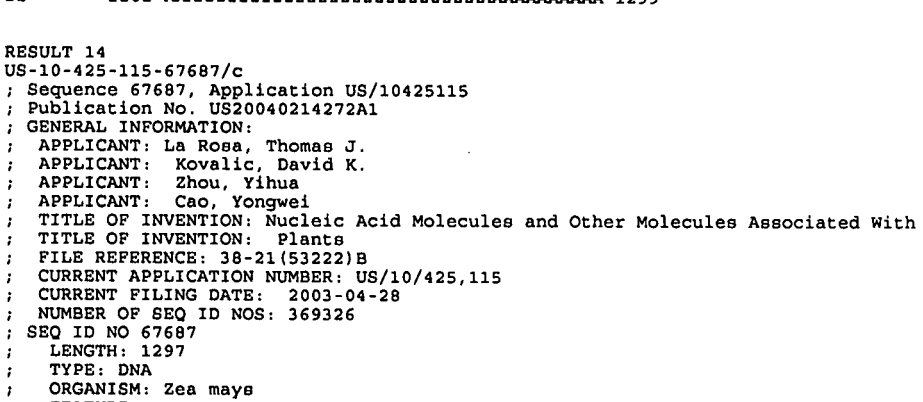
Qy         991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCT 1050
               |||  |      |   ||  |    ||| ||  ||  |  |  |  ||| ||
Db         901 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 960

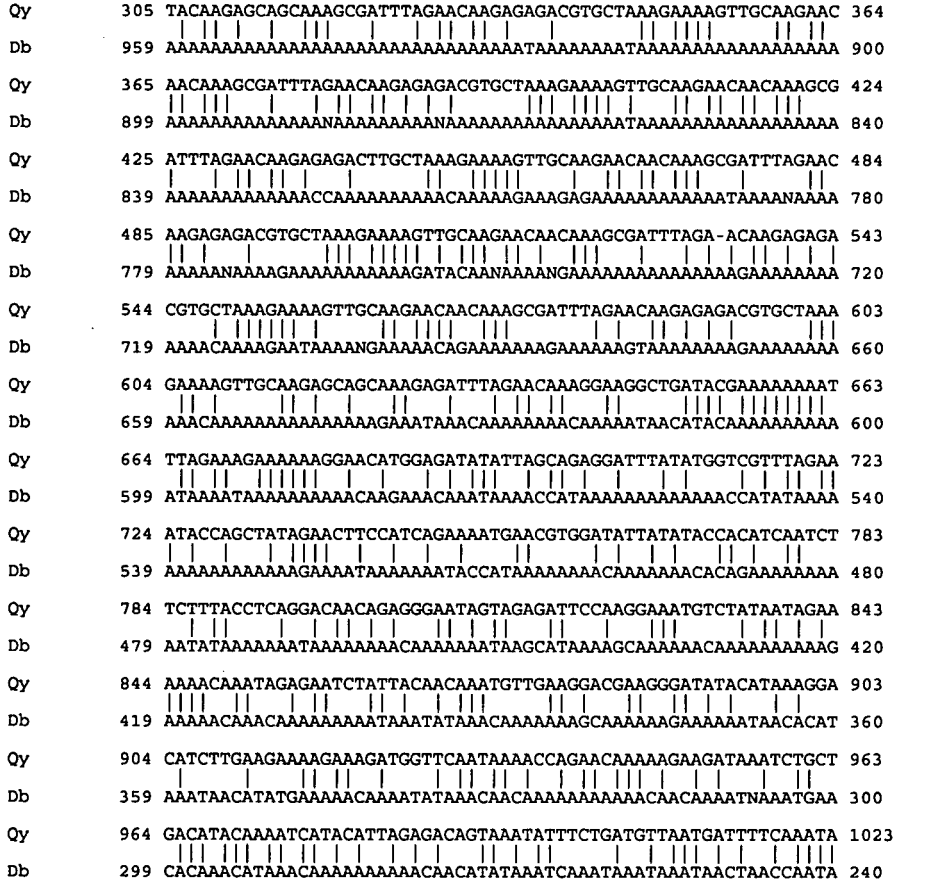
Qy        1051 GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAG 1110
               || |  |  |   |  || | |  || || || |  |  || |   | |  ||   ||
```



RESULT 14
US-10-425-115-67687/c
; Sequence 67687, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67687
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1297)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_161728C.1
US-10-425-115-67687

Query Match 10.2%; Score 151.6; DB 8; Length 1297;
Best Local Similarity 47.1%; Pred. No. 3e-12;
Matches 480; Conservative 0; Mismatches 539; Indels 1; Gaps 1;

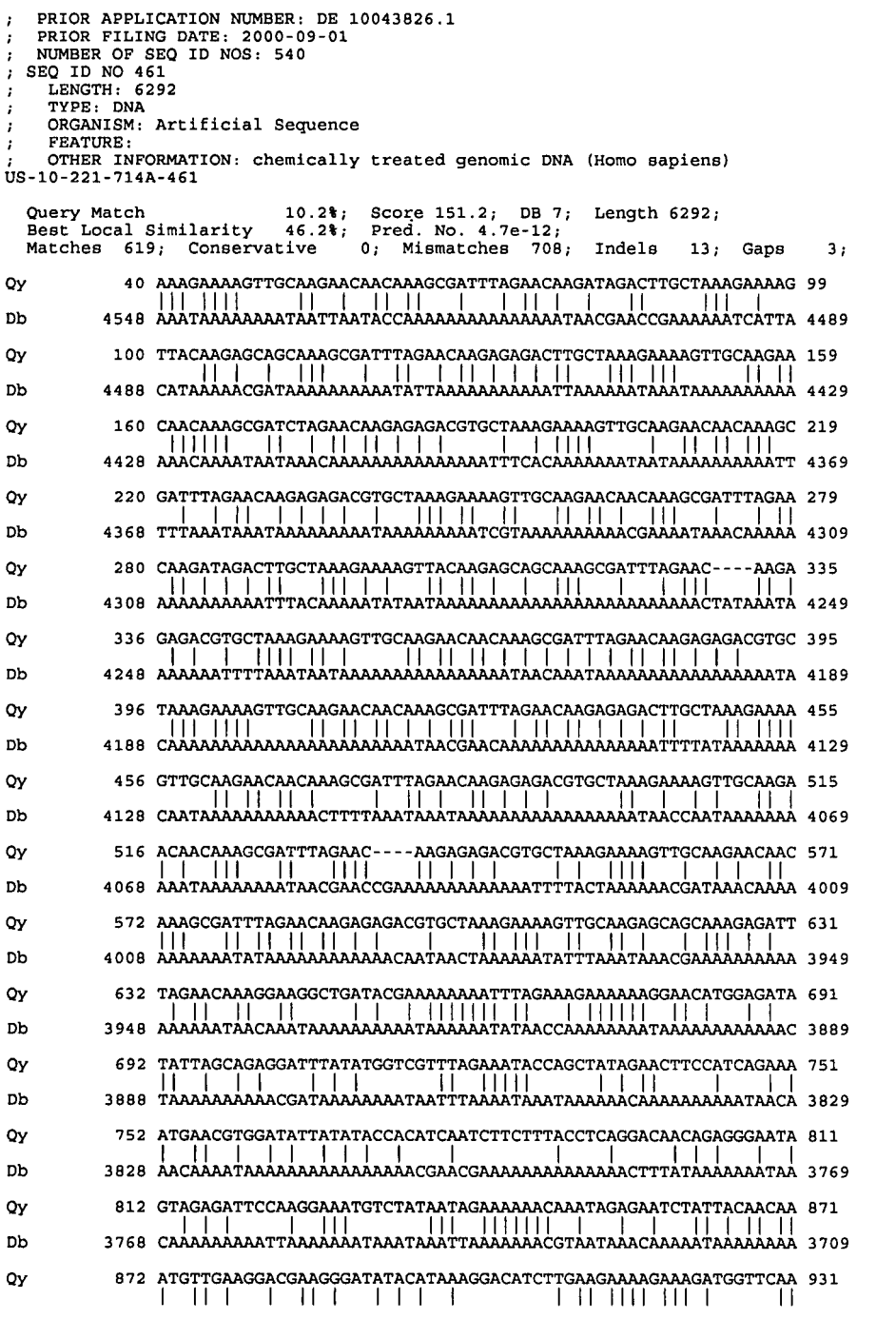


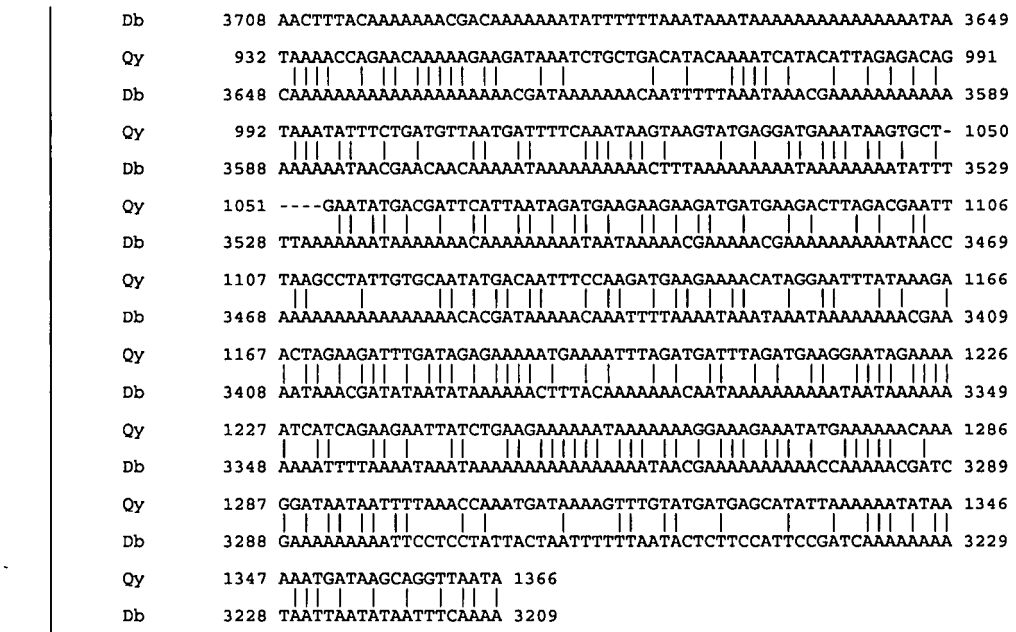


RESULT 15
US-10-221-714A-461/c
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30

```
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
;   LENGTH: 6292
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

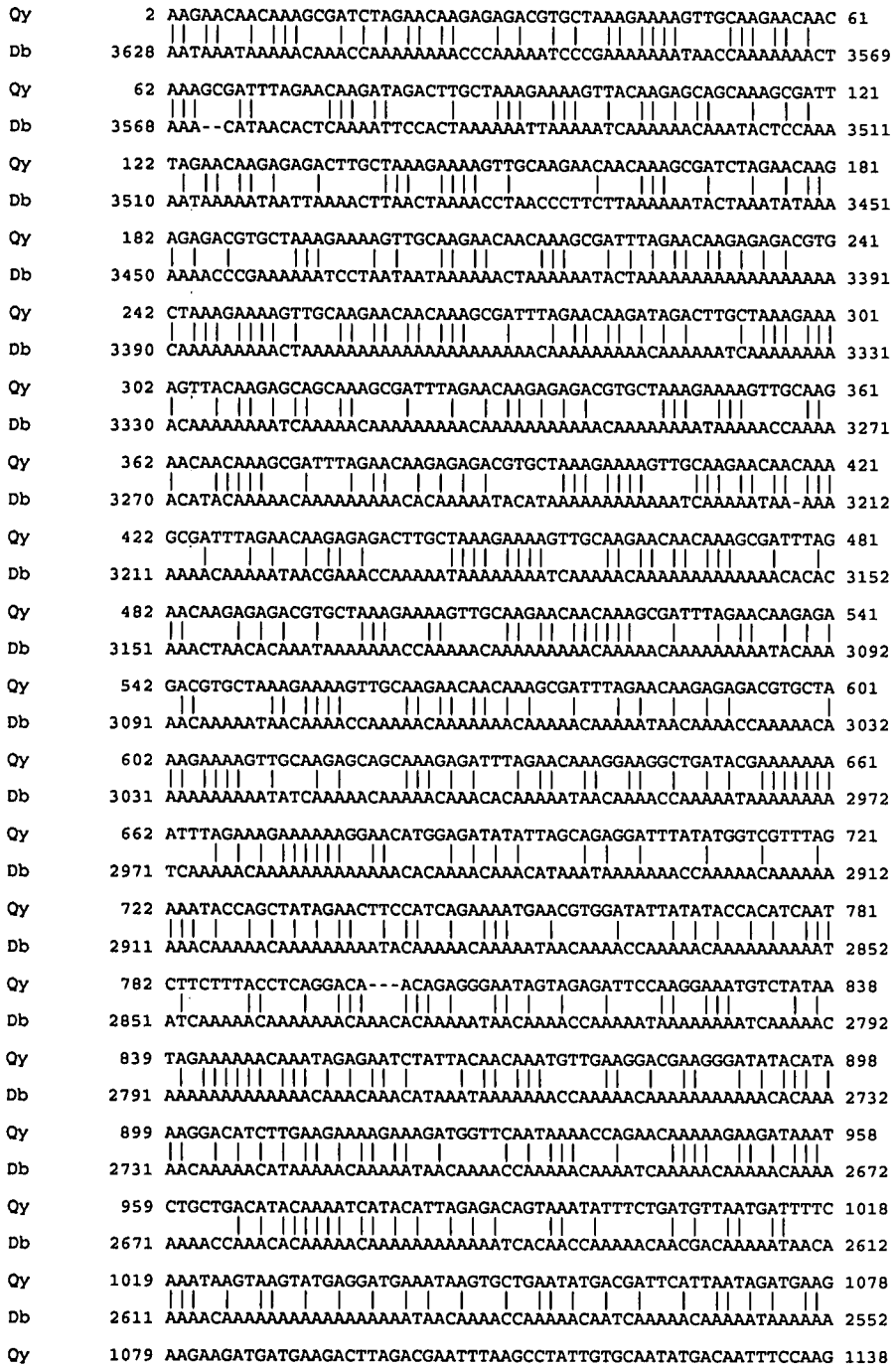
  Query Match             10.2%;  Score 151.2;  DB 7;  Length 6292;
  Best Local Similarity   46.2%;  Pred. No. 4.7e-12;
  Matches  619;  Conservative    0;  Mismatches  708;  Indels   13;  Gaps    3;
```

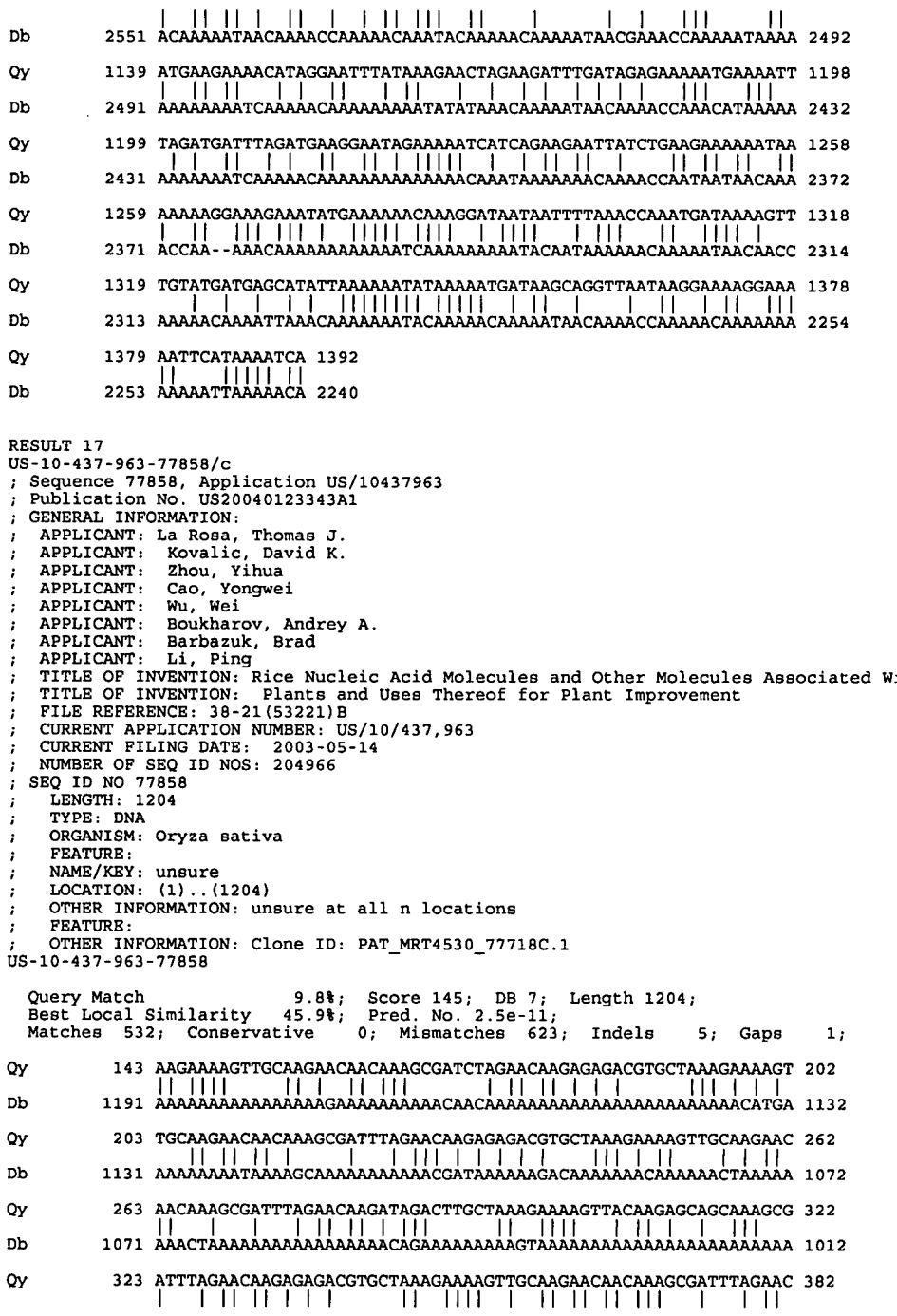




```
RESULT 16
US-10-221-714A-330/c
; Sequence 330, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 330
;   LENGTH: 6767
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-330

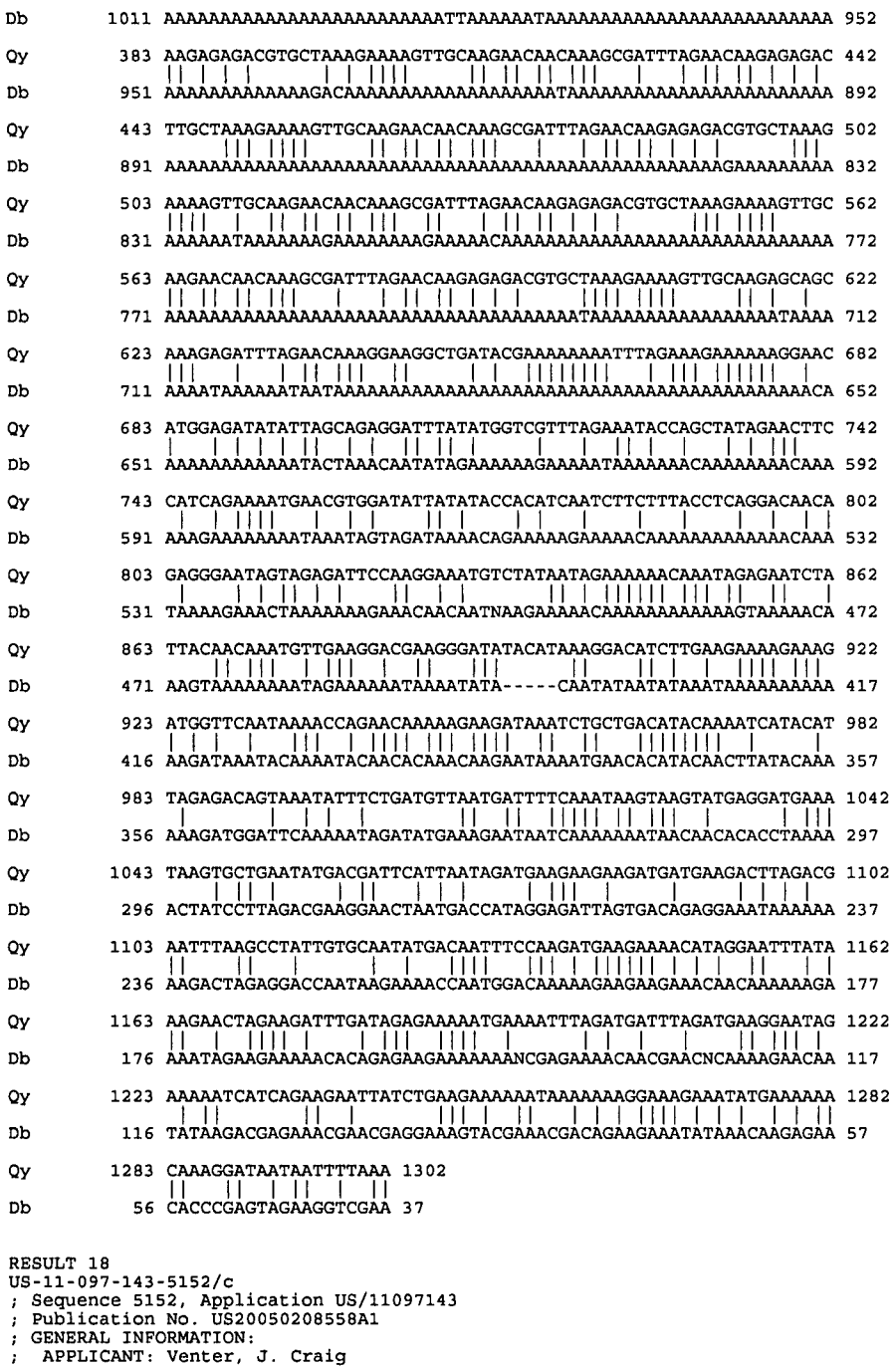
  Query Match             10.1%;  Score 149.2;  DB 7;  Length 6767;
  Best Local Similarity   46.1%;  Pred. No. 9.1e-12;
  Matches  643;  Conservative    0;  Mismatches  743;  Indels    8;  Gaps    4;
```





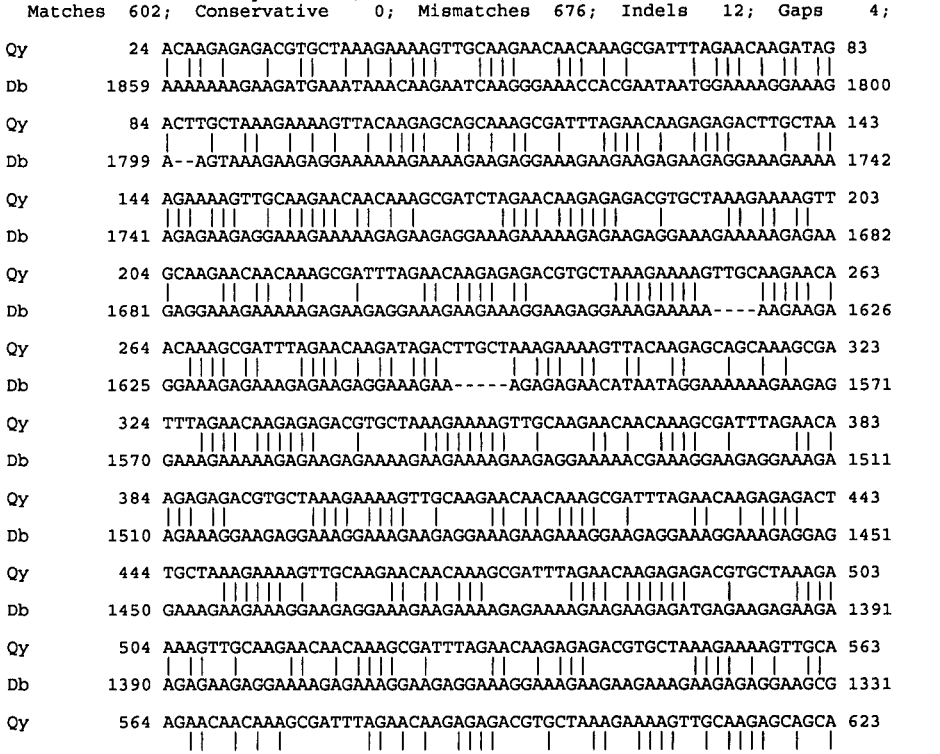
RESULT 17
US-10-437-963-77858/c
; Sequence 77858, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 77858
; LENGTH: 1204
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1204)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_77718C.1
US-10-437-963-77858

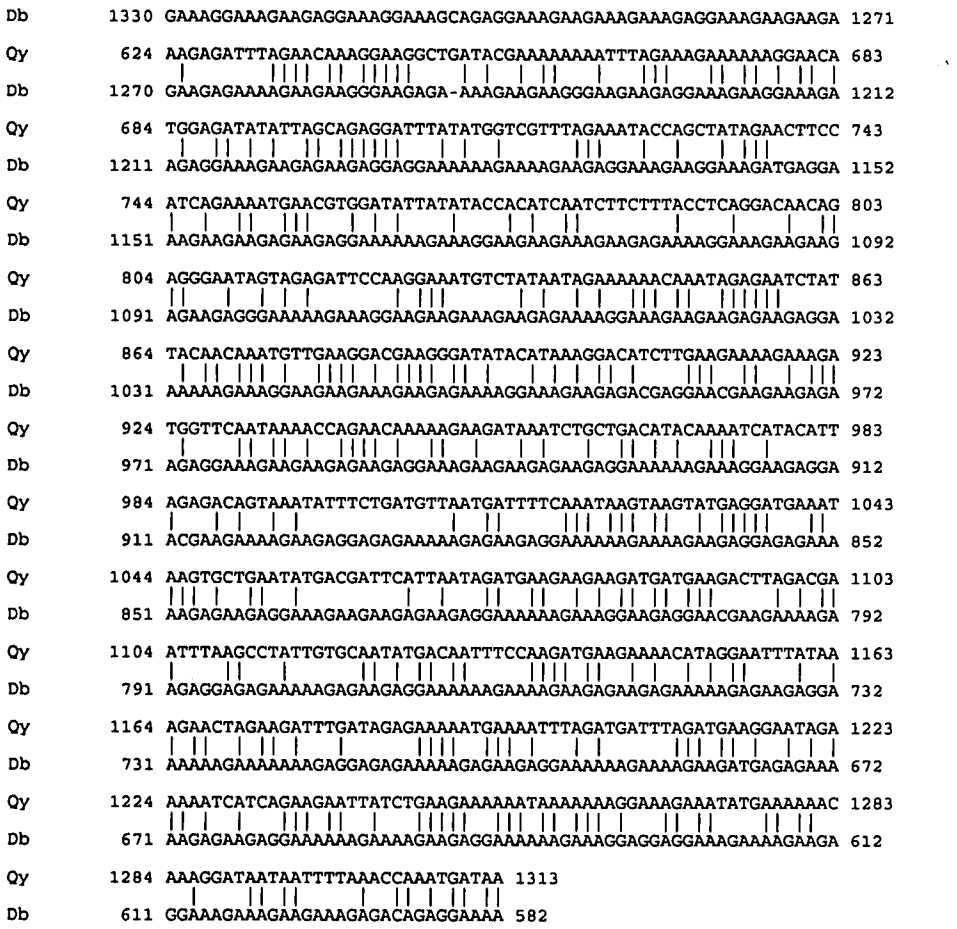
Query Match 9.8%; Score 145; DB 7; Length 1204;
Best Local Similarity 45.9%; Pred. No. 2.5e-11;
Matches 532; Conservative 0; Mismatches 623; Indels 5; Gaps 1;



RESULT 18
US-11-097-143-5152/c
; Sequence 5152, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5152
; LENGTH: 3662
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-5152

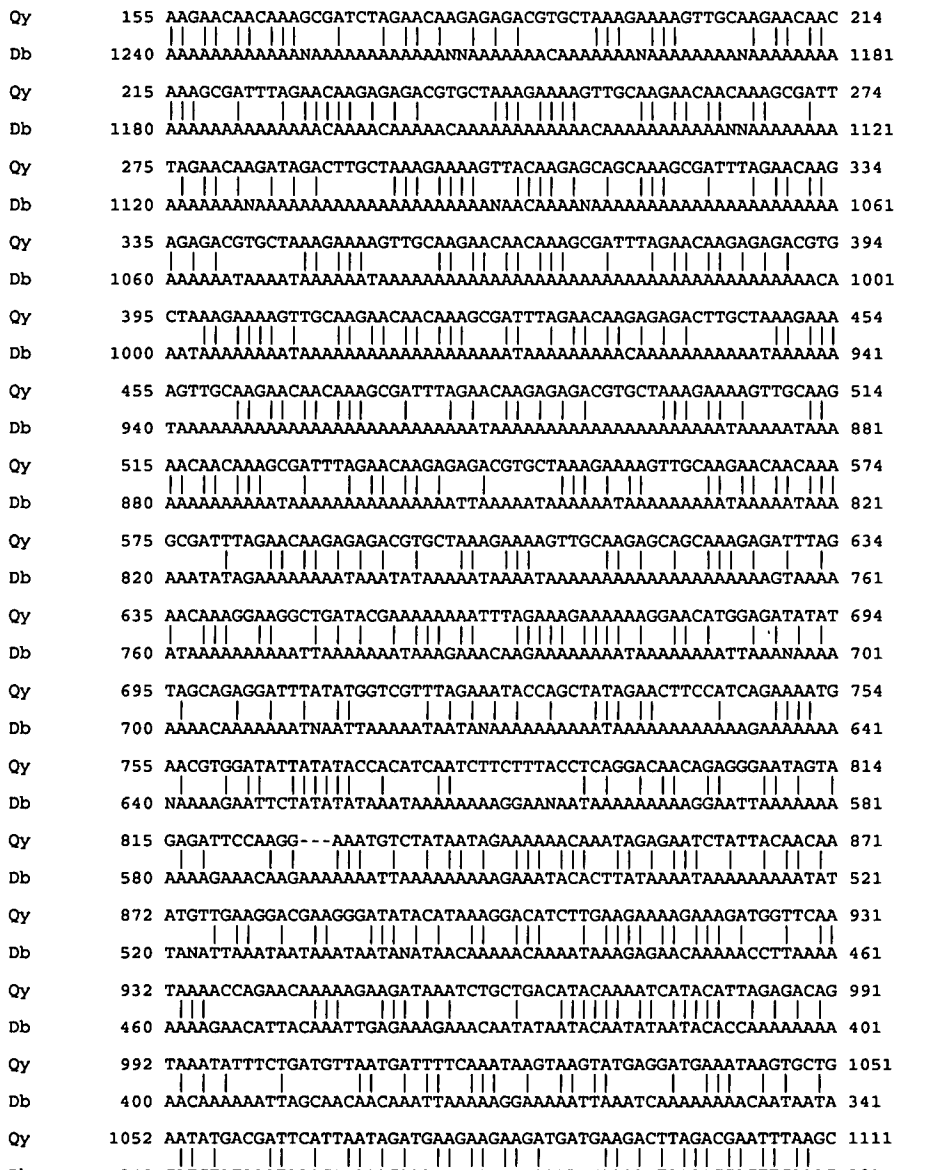
Query Match 9.7%; Score 144.4; DB 10; Length 3662;
Best Local Similarity 46.7%; Pred. No. 3.8e-11;
Matches 602; Conservative 0; Mismatches 676; Indels 12; Gaps 4;





RESULT 19
US-10-425-115-172717/c
; Sequence 172717, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172717
; LENGTH: 1243
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1243)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717

Query Match 9.7%; Score 144.2; DB 8; Length 1243;
Best Local Similarity 45.5%; Pred. No. 3.3e-11;
Matches 516; Conservative 0; Mismatches 616; Indels 3; Gaps 1;



```
Qy       1112 CTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAG 1171
Db        280 AGAACAACCAAAGAGCAAGTAAAATAAAAATATAGAAGGATAAAGAACTAAGTATTCTAG 221

Qy       1172 AAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCAT 1231
Db        220 AAAAAAGAACAAATAAAAATAAAAACAAGAAACAATTATACCACATAAGAAAGAACGAGC 161

Qy       1232 CAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAACAAA 1286
Db        160 AAACAAGAAGAAGAAAGAAACGNACGAACGAGAGAAAATAAAAAGAAAAAGAGAA 106


RESULT 20
US-10-425-115-16756/c
; Sequence 16756, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 16756
;   LENGTH: 1081
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1081)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_115279C.1
US-10-425-115-16756

  Query Match             9.5%;  Score 140.8;  DB 8;  Length 1081;
  Best Local Similarity   47.2%;  Pred. No. 9.5e-11;
  Matches  412;  Conservative    0;  Mismatches  461;  Indels    0;  Gaps    0;

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
Db       1067 AAAAAAAGGAATAAAGAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 1008

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
Db       1007 TAAAAGTAAAAGAAACAAAAAAATAAAAAAAAACACACAAAAGAAAAAAACAAATAAAAA 948

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db        947 AAAATAAAAGAAAAAAGAAAAAAAAAGAAAAAAAAAAAAGTAAAAAAAAAAAAAACAAAA 888

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db        887 AAAAAAACAAAAAAAGANAAAAAAACCAAAATAAACAAAAGAAAATAAAAACCAAAAAAA 828

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
Db        827 AAAAAAAAAAAAAAAAAAAAAAACAAAAAAGAAAAAAAAAAAAAAAAAGCAAAAAAAAAC 768

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
Db        767 AAAAAAAAAAAAAAAATAAATAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAA 708

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
Db        707 AAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAACAAAAAAAAAGAAAAAA 648

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
```

```
Db        647 AAAAACAATAAAATAAAATAATATAAAAAAAAACANAAAAACAAACAAAAATAAAGAATA 588

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
Db        587 AAAAATAAAACAAAAAGNNAAAACAGNAAAAAAACANATAAACAAAAAAAAAAATAGAAT 528

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
Db        527 AAAAAAGAAANAAAAAATACAANAAATAAGAAAAAACATAATAAACAAAAAAAAAACAAA 468

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
Db        467 AAAAATATGAAAAAATAACAAAAAAACACNAAAAAAAACAAAAAAAAAAAAAAAAACAA 408

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
Db        407 TATAATAAGAAAAAAAAAAAAAAAAAAAAAGATATAAAAAAATAAAAGAAGAAAAAAAAA 348

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
Db        347 AAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAACAAAAAAAAAAATAAAAA 288

Qy       1330 CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db        287 AAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAACAAAAAAACAGAAAAA 228

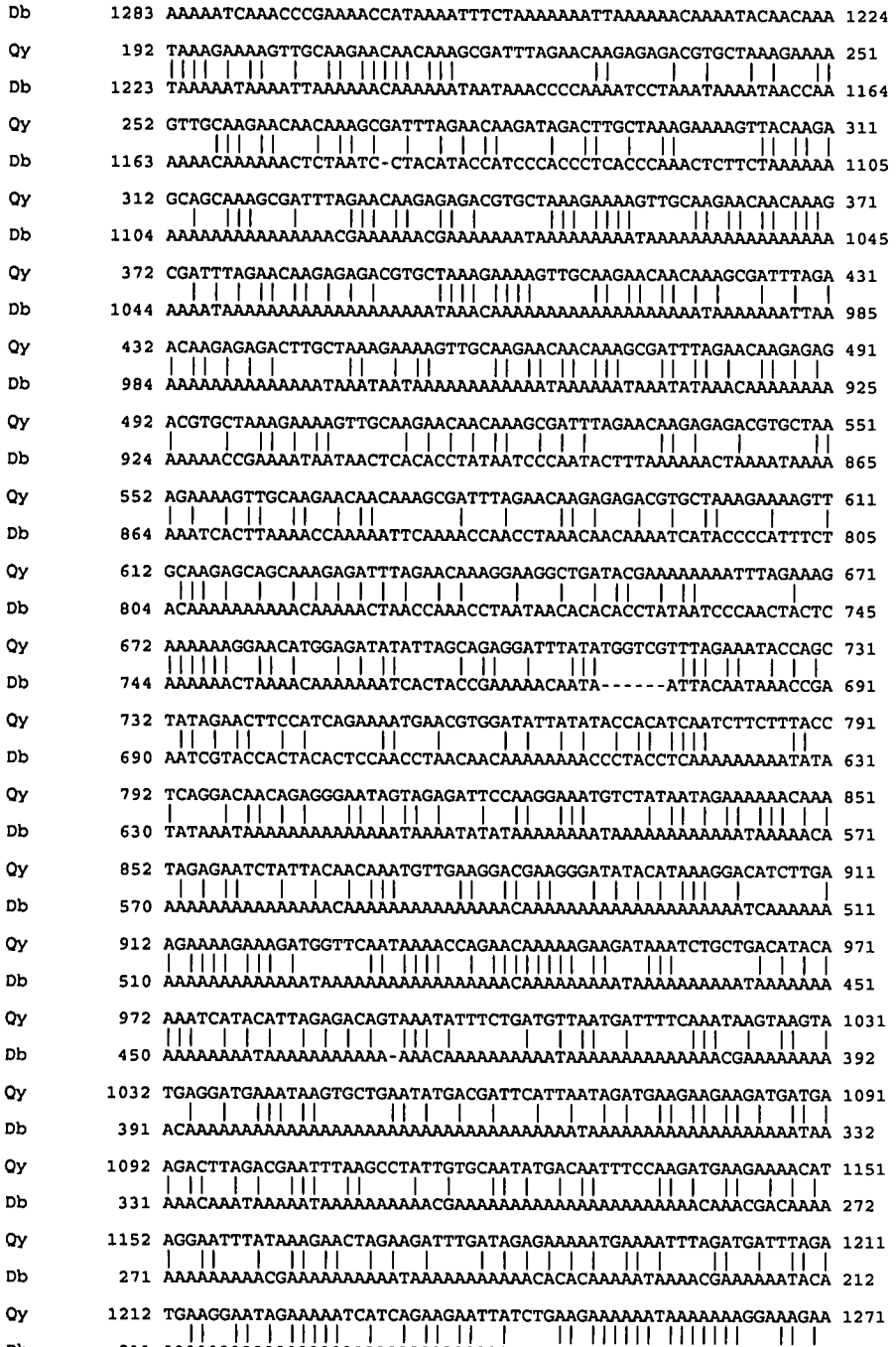
Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAA 1422
Db        227 GACTCATAACAAACATACAAAAAACAAAAAGCA 195

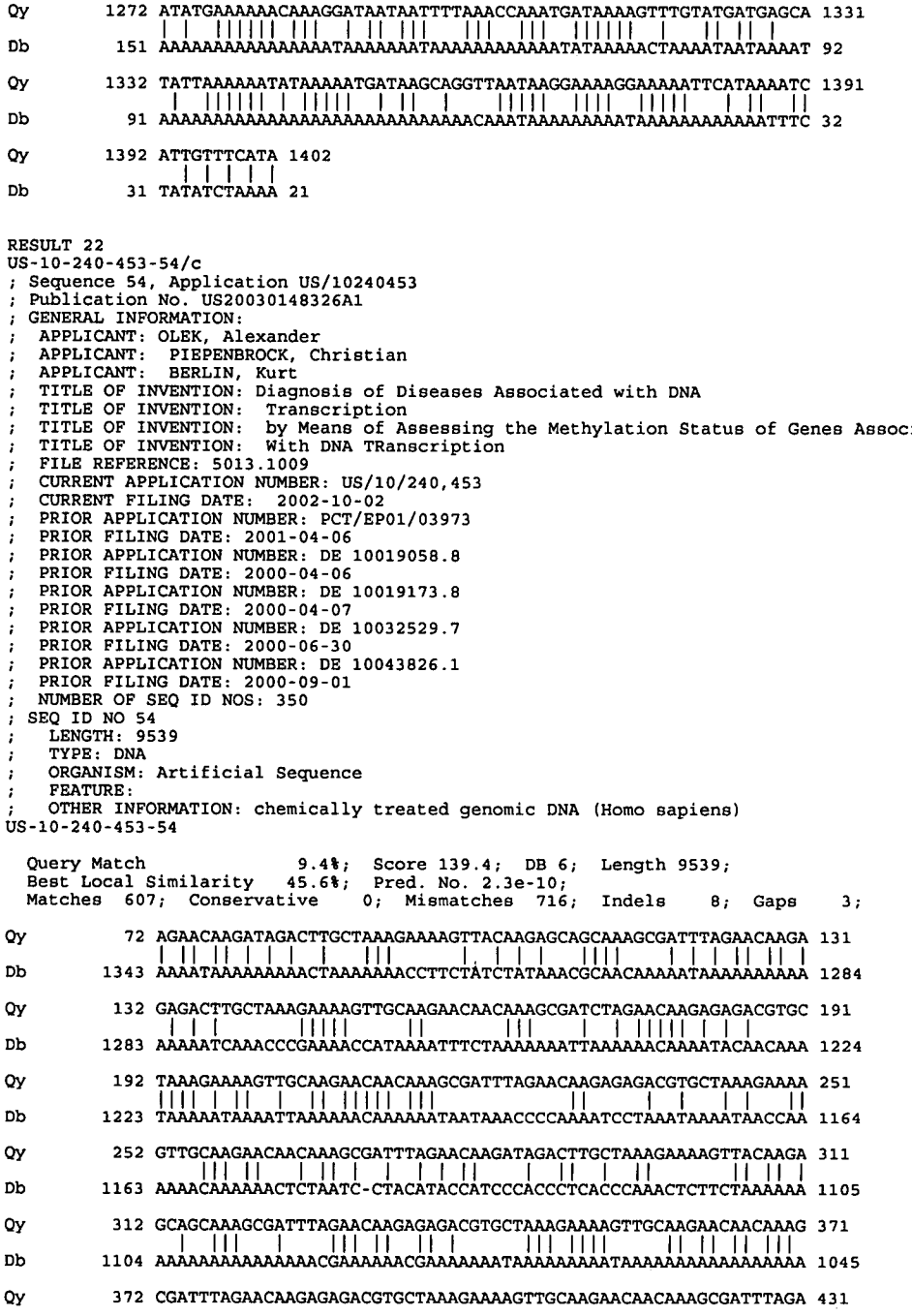

RESULT 21
US-10-239-676-52/c
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 52
;   LENGTH: 9539
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-52

  Query Match             9.4%;  Score 139.4;  DB 5;  Length 9539;
  Best Local Similarity   45.6%;  Pred. No. 2.3e-10;
  Matches  607;  Conservative    0;  Mismatches  716;  Indels    8;  Gaps    3;

Qy         72 AGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGA 131
Db       1343 AAAATAAAAAAAAACTAAAAAAACCTTCTATCTATAAACGCAACAAAAATAAAAAAAAAA 1284

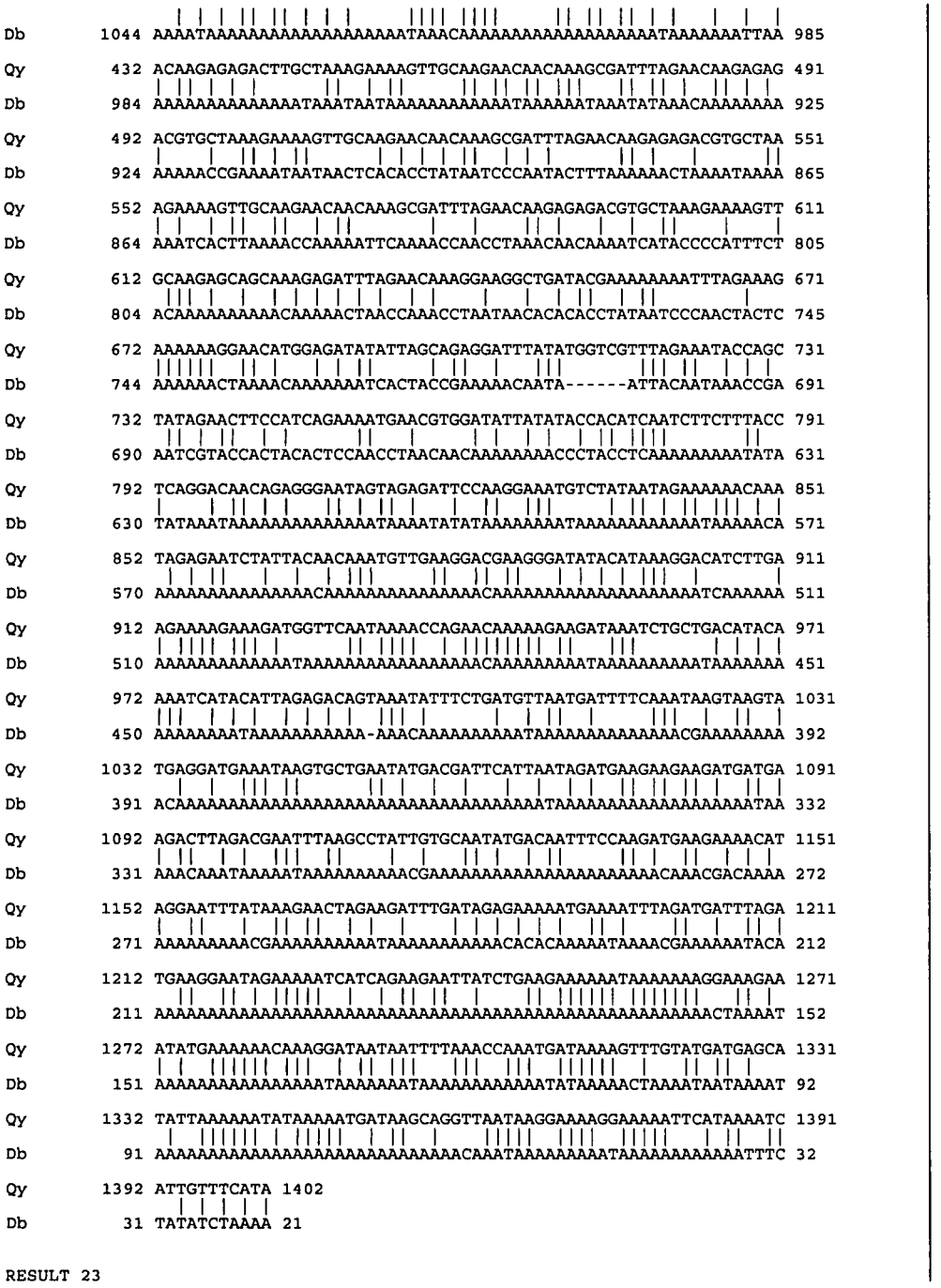
Qy        132 GAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGC 191
```





RESULT 22
US-10-240-453-54/c
; Sequence 54, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associate
; TITLE OF INVENTION: With DNA TRanscription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 54
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-54

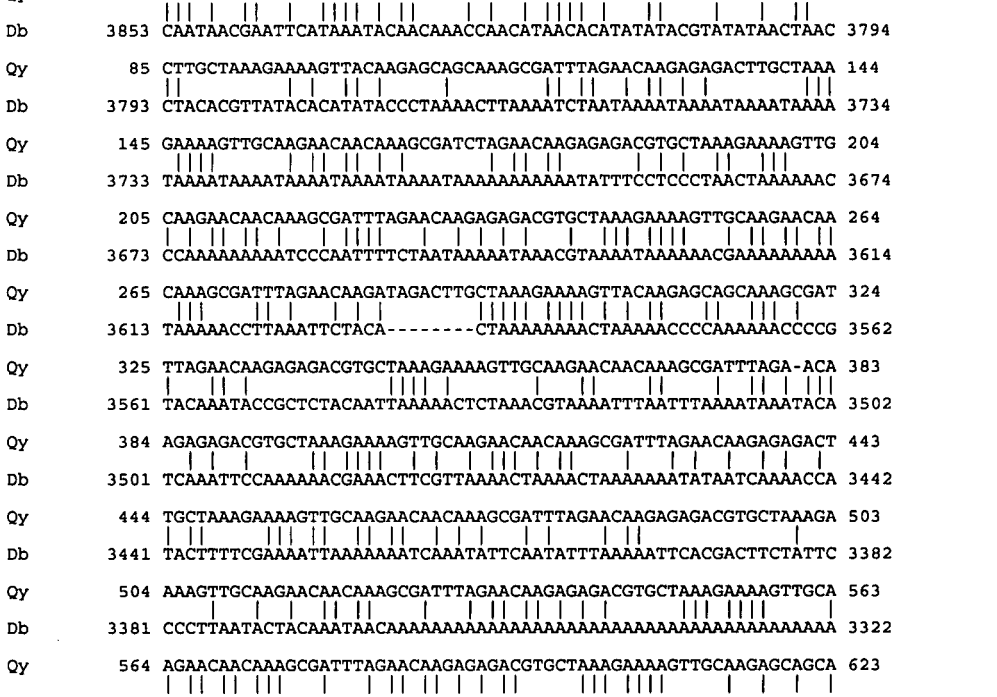
Query Match 9.4%; Score 139.4; DB 6; Length 9539;
Best Local Similarity 45.6%; Pred. No. 2.3e-10;
Matches 607; Conservative 0; Mismatches 716; Indels 8; Gaps 3;

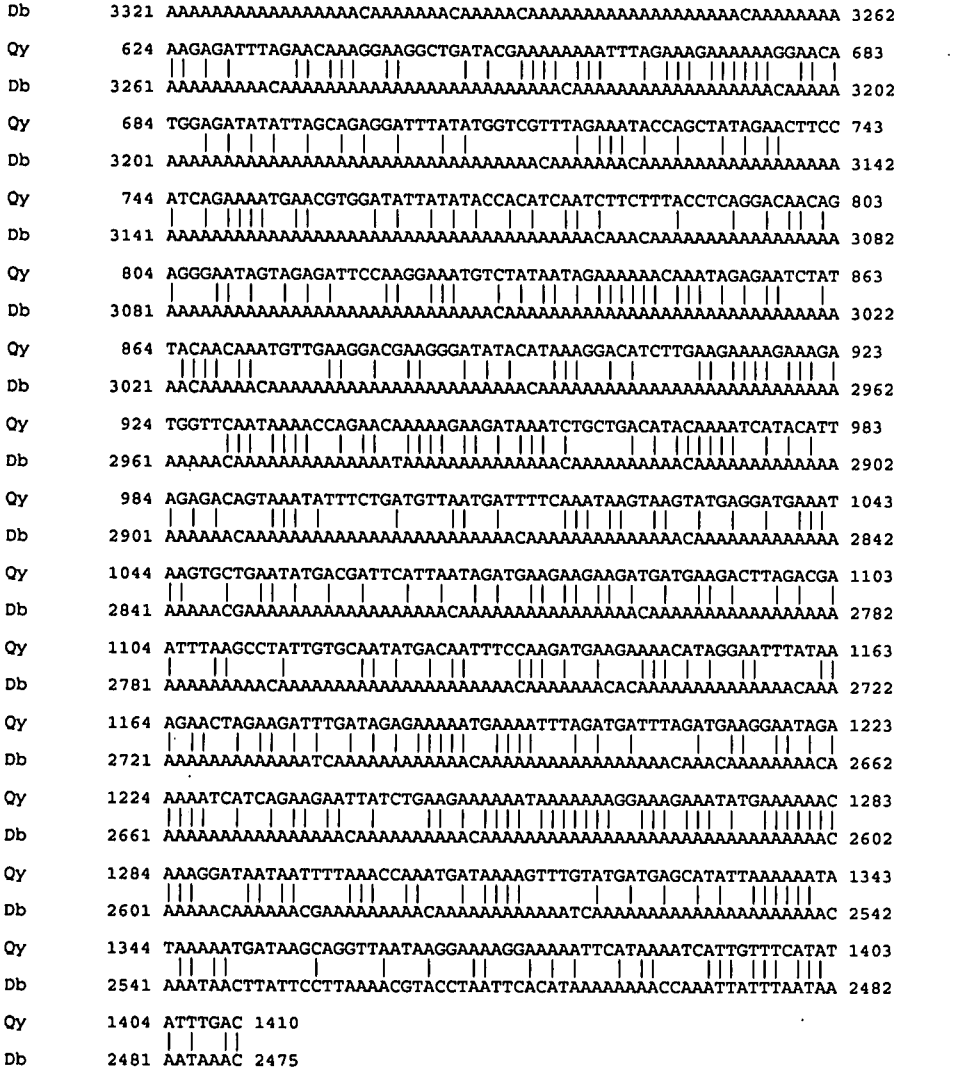


RESULT 23

US-10-311-455-1670/c
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1670
; LENGTH: 6668
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1936
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

Query Match 9.2%; Score 136.2; DB 6; Length 6668;
Best Local Similarity 44.7%; Pred. No. 6.2e-10;
Matches 620; Conservative 0; Mismatches 758; Indels 9; Gaps 2;





RESULT 24
US-10-363-345A-2179/c
; Sequence 2179, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712

; SEQ ID NO 2179
; LENGTH: 778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CpG-island No: 2179
US-10-363-345A-2179

Query Match 8.9%; Score 131.2; DB 8; Length 778;
Best Local Similarity 48.4%; Pred. No. 2e-09;
Matches 364; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

RESULT 25
US-10-363-345A-2180
; Sequence 2180, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock

```
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Method for determining the degree of methylation of defined
;  TITLE OF INVENTION:  cytosines in genomic DNA in the sequence context of 5'-CpG-3
;  FILE REFERENCE: E01/1227
;  CURRENT APPLICATION NUMBER: US/10/363,345A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2180
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2180
US-10-363-345A-2180

  Query Match             8.9%;  Score 131.2;  DB 8;  Length 778;
  Best Local Similarity   48.4%;  Pred. No. 2e-09;
  Matches  364;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||    | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db          6 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 65

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
                | |   | |       ||  |||| |  |   | | |      | | ||| |  |
Db         66 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 125

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
              | |  |  ||| | |  ||  |  | ||    |  | |  | |   || |  | |
Db        126 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAAATTATAAAAAAAAAAAAAA 185

Qy        818 ATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  |||     || ||| ||||||  ||| |    |   |  | || ||||   |
Db        186 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTACG 245

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |    ||    ||||
Db        246 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAA 305

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | ||||  | |     | | |  ||| |
Db        306 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 365

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |    ||  ||| | ||| || ||| || ||  |  ||| ||       | |
Db        366 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 425

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   |  || || || || || | |    | |    |  ||    |
Db        426 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTACGAAAAAAAAAT 485

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||           || ||||  || |||   | | |  || | || |
Db        486 TACGAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAA 545

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |   | |||  ||||   | || |   | |  ||  || | |  |  |  | ||
Db        546 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 605

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | ||| | |||  | || ||||
Db        606 AAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAAATAACAAAAAAAAAAAAATT 665

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
              ||||  |||  | |  |      |  |  |  ||  ||||||| | |||||     |
Db        666 ATAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAA 725

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db        726 AAAAAAAAAATTATAAAAAAAAAAAATTACAA 757
```

```
RESULT 26
US-10-363-483A-2179/c
; Sequence 2179, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2179
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2179
US-10-363-483A-2179

  Query Match             8.9%;  Score 131.2;  DB 9;  Length 778;
  Best Local Similarity   48.4%;  Pred. No. 2e-09;
  Matches  364;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||    | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db        773 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 714

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
                | |   | |       ||  |||| |  |   | | |      | | ||| |  |
Db        713 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 654

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
              | |  |  ||| | |  ||  |  | ||    |  | |  | |   || |  | |
Db        653 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAAATTATAAAAAAAAAAAAAA 594

Qy        818 ATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  |||     || ||| ||||||  ||| |    |   |  | || ||||   |
Db        593 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACG 534

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |    ||    ||||
Db        533 AAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAA 474

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | ||||  | |     | | |  ||| |
Db        473 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 414

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |    ||  ||| | ||| || ||| || ||  |  ||| ||       | |
Db        413 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 354

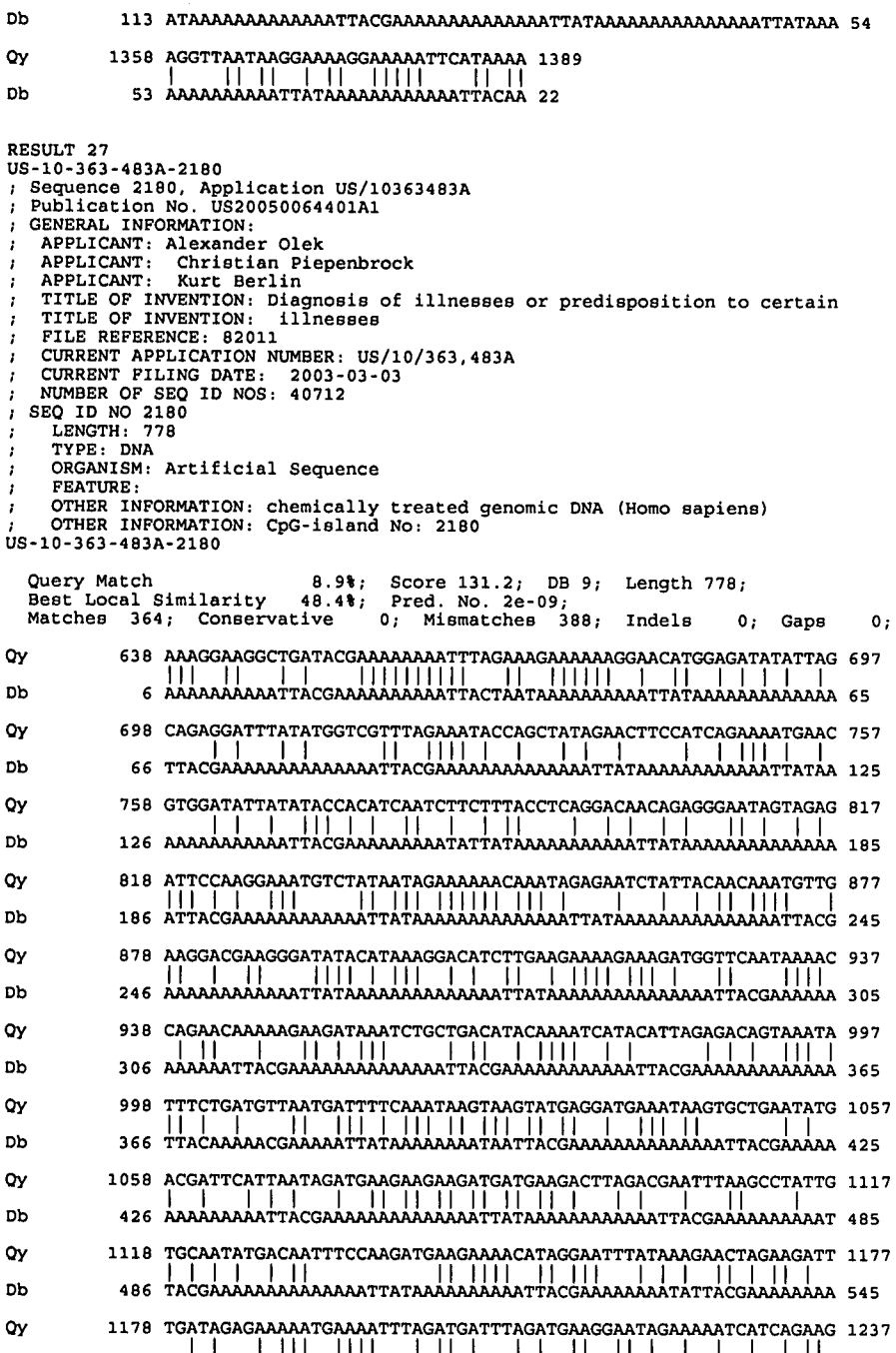
Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   |  || || || || || | |    | |    |  ||    |
Db        353 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAATTACGAAAAAAAAAT 294

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||           || ||||  || |||   | | |  || | || |
Db        293 TACGAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAA 234

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |   | |||  ||||   | || |   | |  ||  || | |  |  |  | ||
Db        233 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 174

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  | ||| | |||  | || ||||
Db        173 AAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAATT 114

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
              ||||  |||  | |  |      |  |  |  ||  ||||||| | |||||     |
```



```
Db        113 ATAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAAATTATAAA 54

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db         53 AAAAAAAAAATTATAAAAAAAAAAAAATTACAA 22


RESULT 27
US-10-363-483A-2180
; Sequence 2180, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
;  TITLE OF INVENTION:  illnesses
;  FILE REFERENCE: 82011
;  CURRENT APPLICATION NUMBER: US/10/363,483A
;  CURRENT FILING DATE:  2003-03-03
;  NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 2180
;   LENGTH: 778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   OTHER INFORMATION: CpG-island No: 2180
US-10-363-483A-2180

  Query Match             8.9%;  Score 131.2;  DB 9;  Length 778;
  Best Local Similarity   48.4%;  Pred. No. 2e-09;
  Matches  364;  Conservative    0;  Mismatches  388;  Indels    0;  Gaps    0;

Qy        638 AAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAG 697
              |||  ||   | |    ||||||||||   ||  ||||||  |  ||  | | | |  |
Db          6 AAAAAAAAAATTACGAAAAAAAAAATTACTAATAAAAAAAAAATTATAAAAAAAAAAAAA 65

Qy        698 CAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAAC 757
                | |   | |      ||  |||| |  |   | |  |     |  | |||  |  |
Db         66 TTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAATTATAA 125

Qy        758 GTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAG 817
              | |  |  ||| | |  ||  | | ||    |  | |  | |   || | |  |
Db        126 AAAAAAAAAAATTACGAAAAAAAAATATTATAAAAAAAAAAATTATAAAAAAAAAAAAAA 185

Qy        818 ATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTG 877
              ||| | |  |||     || ||| |||||| ||| |   |   | | || ||||   |
Db        186 ATTACGAAAAAAAAAAAATTATAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTACG 245

Qy        878 AAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAAC 937
              ||  |  ||    |||| | |||  | |  ||  | |||| ||| |   ||    ||||
Db        246 AAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTACGAAAAAA 305

Qy        938 CAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATA 997
              | ||   |  || | |||     | ||  | ||||  | |       | | |  ||| |
Db        306 AAAAAATTACGAAAAAAAAAAAAAATTACGAAAAAAAAAAAATTACGAAAAAAAAAAAAA 365

Qy        998 TTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATG 1057
              || |  |   ||  ||| | ||| || ||| || ||  |  ||| ||      | |
Db        366 TTACAAAAACGAAAAATTATAAAAAAAATAATTACGAAAAAAAAAAAAAATTACGAAAAA 425

Qy       1058 ACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTG 1117
              |  |   | | |   || || || || || || |   | |   |   ||    |
Db        426 AAAAAAAAATTACGAAAAAAAAAAAAAAATTATAAAAAAAAAAAAATTACGAAAAAAAAAT 485

Qy       1118 TGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATT 1177
              | | | |  | ||           || ||||  || |||   | | |  || | || |
Db        486 TACGAAAAAAAAAAAAAAATTATAAAAAAAAAATTACGAAAAAAAATATTACGAAAAAAA 545

Qy       1178 TGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAG 1237
                | |   | |||  ||||  | || |  | |  ||  || | |  |  | |  | ||
Db        546 AAAAAATTATAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAA 605

Qy       1238 AATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATT 1297
              || |  |  || |||||| |||  | | ||| ||| |  |  || | |||  | || ||||
Db        606 AAATTACGAAAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATAACAAAAAAAAAAAAAATT 665

Qy       1298 TTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGC 1357
              ||||  |||  | |  |       |  |  |  ||  ||||||| | |||||     |
Db        666 ATAAAAAAAAAAAAATTACGAAAAAAAAAAAAAATTATAAAAAAAAAAAAAAAATTATAAA 725

Qy       1358 AGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |    || ||  | ||  |||||    || ||
Db        726 AAAAAAAAAATTATAAAAAAAAAAAAAATTACAA 757


RESULT 28
US-10-221-714A-409/c
; Sequence 409, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 409
;   LENGTH: 7442
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1792, 4861..4862
;   OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-409

  Query Match             8.8%;  Score 129.8;  DB 7;  Length 7442;
  Best Local Similarity   46.7%;  Pred. No. 5e-09;
  Matches  564;  Conservative    0;  Mismatches  627;  Indels   18;  Gaps    4;

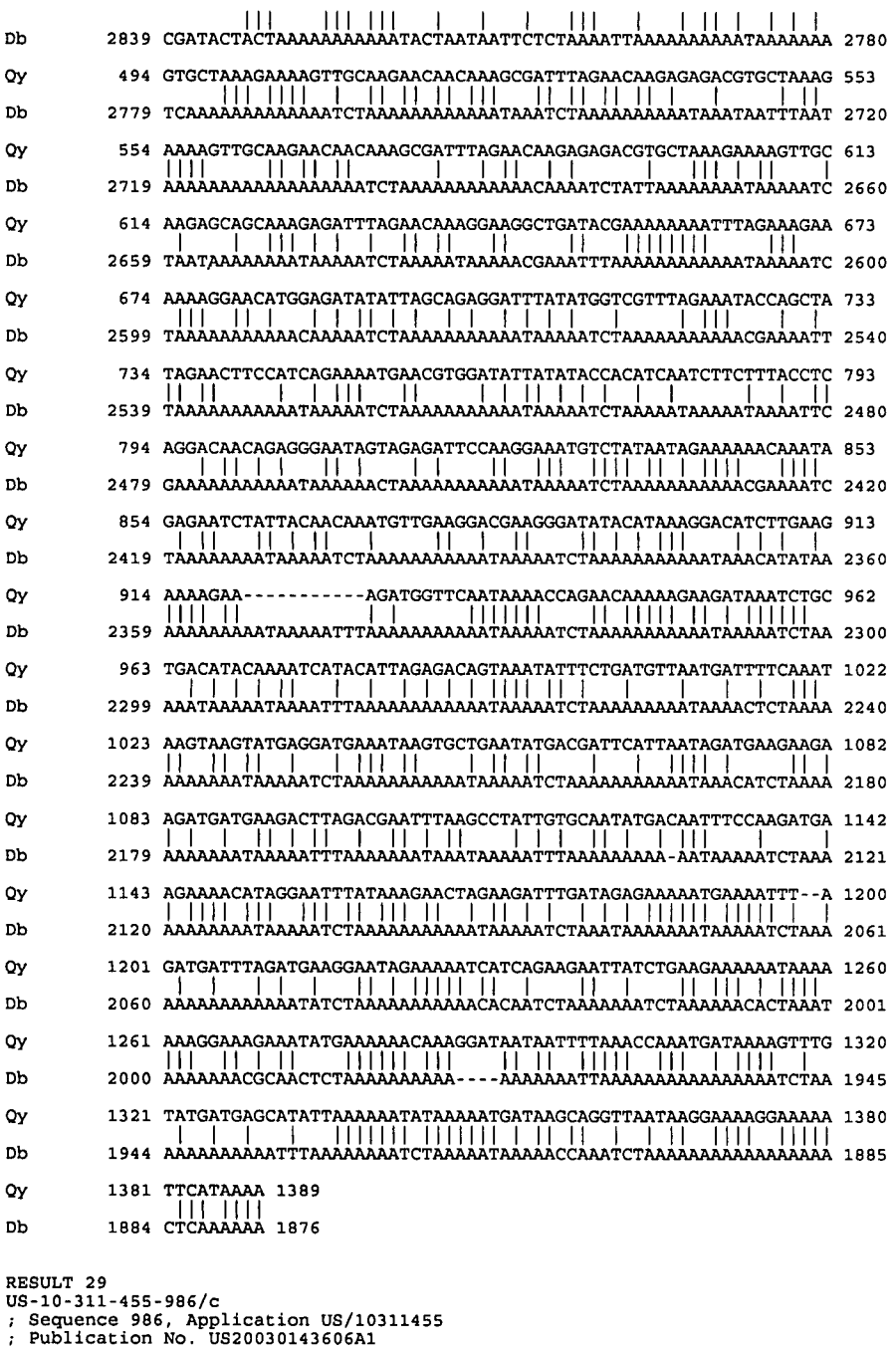
Qy        194 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 253
              || ||   |   |  | || |||   ||  |  || |      |    | ||| ||||
Db       3079 AAAAACCCTCAAACACAAAAAAAAAATATAAAACACTATCCCAATACACAAAAAAAAAAA 3020

Qy        254 TGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGC 313
                ||    || ||| |||     || |  |  |  |  ||| |||| | || |
Db       3019 CTAAAATTTAAAAAAACGAAACTTAAAATAAAAACCCTCAAAAATAAAATAAAAAAAACA 2960

Qy        314 AGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 373
              |||| | ||       |  | |  |       ||  ||       ||   |||| | |
Db       2959 CCCAAACCTATCCCACTCCTAACATCCCAAAACAACTAATACAATAATTCCAACCTAACA 2900

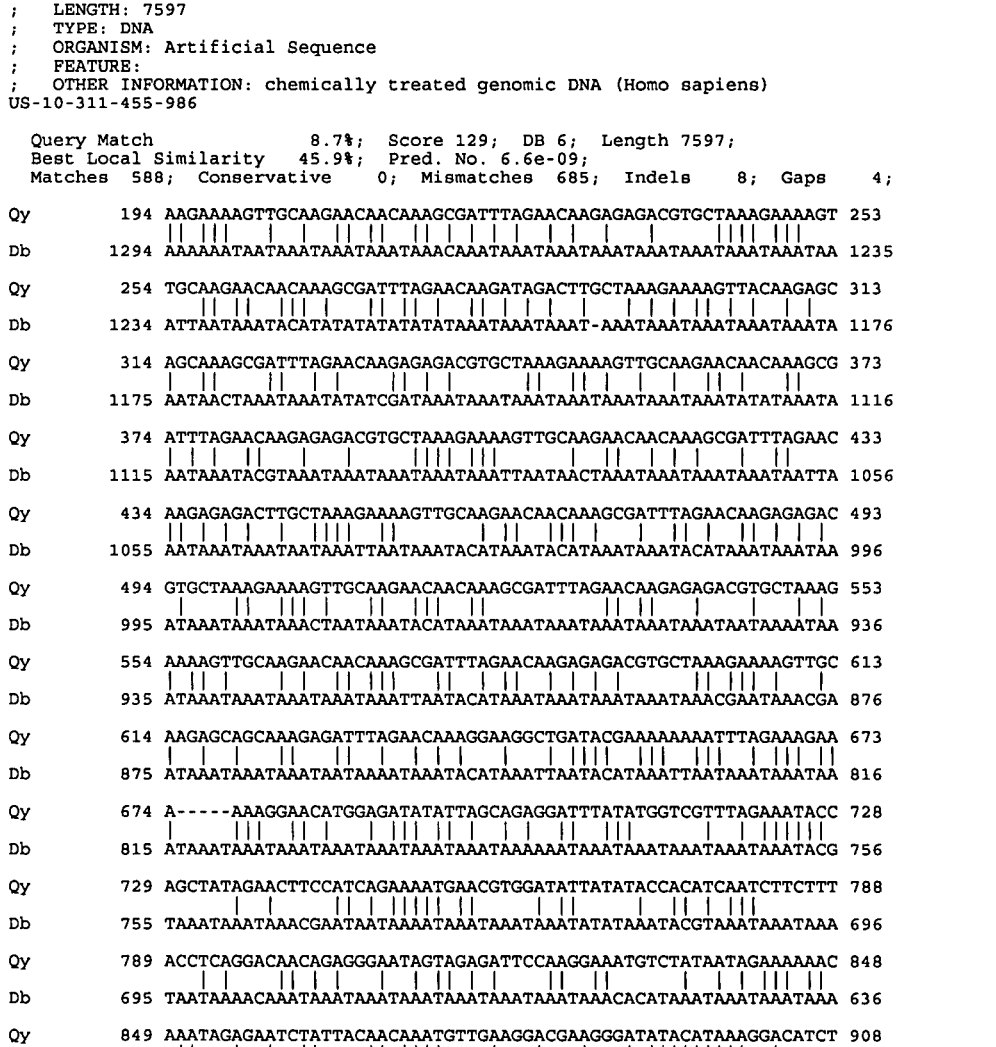
Qy        374 ATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 433
              |      | || | | |     |||  ||       || |     |||       | ||
Db       2899 AAACTATAAAAAAAAAAAAAATATAATAAATCCGATAAAAACTTCGAAAAACTAAAAAAAA 2840

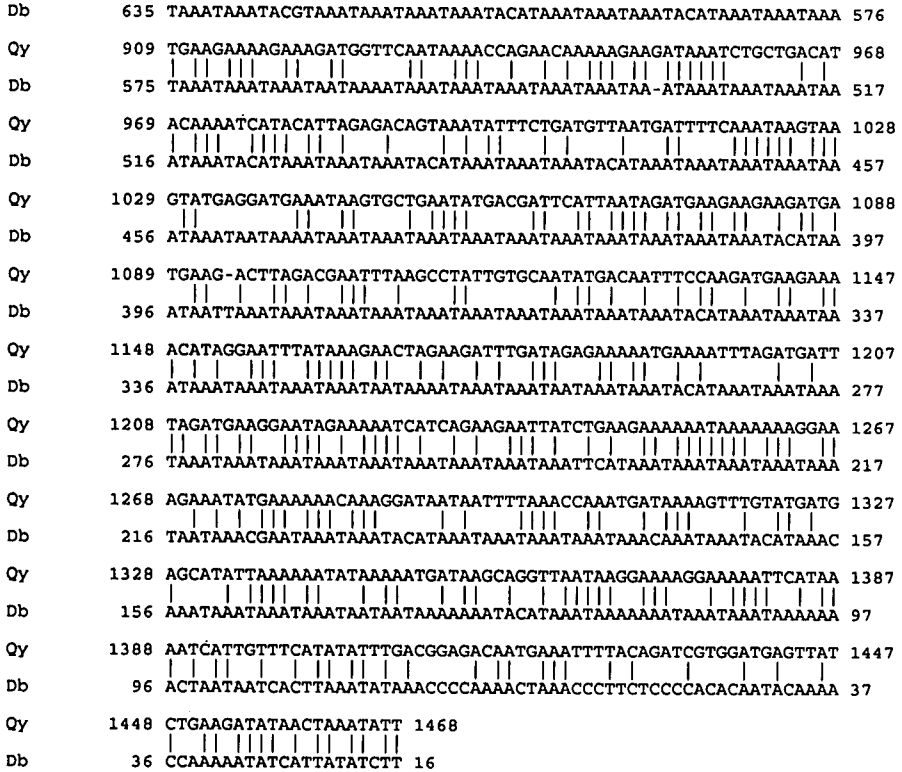
Qy        434 AAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 493
```



RESULT 29
US-10-311-455-986/c
; Sequence 986, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

Query Match 8.7%; Score 129; DB 6; Length 7597;
Best Local Similarity 45.9%; Pred. No. 6.6e-09;
Matches 588; Conservative 0; Mismatches 685; Indels 8; Gaps 4;





```
Db        635 TAAATAAATACGTAAATAAATAAATAAATACATAAATAAATAAATACATAAATAAATAAA 576

Qy        909 TGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACAT 968

Db        575 TAAATAAATAAATAATAAAATAAATAAATAAATAAATAAATAA-ATAAATAAATAAATAA 517

Qy        969 ACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAA 1028

Db        516 ATAAATACATAAATAAATAAATACATAAATAAATAAATACATAAATAAATAAATAAATAA 457

Qy       1029 GTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGA 1088

Db        456 ATAAATAATAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAA 397

Qy       1089 TGAAG-ACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAA 1147

Db        396 ATAATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAA 337

Qy       1148 ACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATT 1207

Db        336 ATAAATAAATAAATAAATAATAAAATAAATAAATAATAAATACATAAATAAATAAA 277

Qy       1208 TAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAA 1267

Db        276 TAAATAAATAAATAAATAAATAAATAAATAAATAAATTCATAAATAAATAAATAAATAAA 217

Qy       1268 AGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATG 1327

Db        216 TAATAAACGAATAAATAAATACATAAATAAATAAATAAATAAACAAATAAATACATAAAC 157

Qy       1328 AGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAA 1387

Db        156 AAATAAATAAATAAATAATAATAAAAAAATACATAAATAAAAAAATAAATAAATAAAAAA 97

Qy       1388 AATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTAT 1447

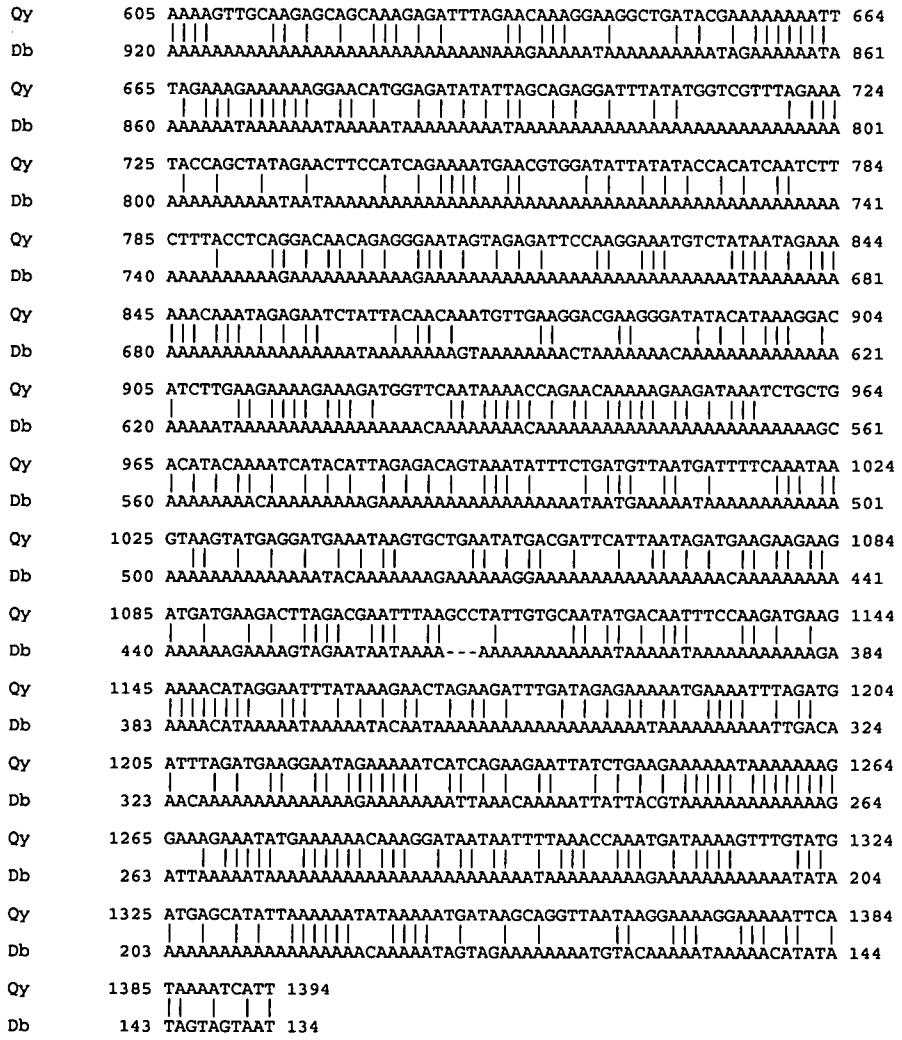
Db         96 ACTAATAATCACTTAAATATAAACCCCAAAACTAAACCCTTCTCCCCACACAATACAAAA 37

Qy       1448 CTGAAGATATAACTAAATATT 1468

Db         36 CCAAAAATATCATTATATCTT 16
```

RESULT 30
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match 8.7%; Score 128.2; DB 8; Length 921;
Best Local Similarity 48.5%; Pred. No. 5.5e-09;
Matches 383; Conservative 0; Mismatches 404; Indels 3; Gaps 1;



```
Qy        605 AAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATT 664

Db        920 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 861

Qy        665 TAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAA 724

Db        860 AAAAAATAAAAAAATAAAAATAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 801

Qy        725 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 784

Db        800 AAAAAAAAAATAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 741

Qy        785 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAA 844

Db        740 AAAAAAAAAAGAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 681

Qy        845 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 904

Db        680 AAAAAAAAAAAAAAAAAATAAAAAAAAGTAAAAAAAACTAAAAAAACAAAAAAAAAAAAA 621

Qy        905 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 964

Db        620 AAAAATAAAAAAAAAAAAAAAAAAACAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAGC 561

Qy        965 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 1024

Db        560 AAAAAAAACAAAAAAAAAGAAAAAAAAAAAAAAAAAAATAATGAAAAATAAAAAAAAAAA 501

Qy       1025 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 1084

Db        500 AAAAAAAAAAAAAAATACAAAAAAAGAAAAAAGGAAAAAAAAAAAAAAAAACAAAAAAAA 441

Qy       1085 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 1144

Db        440 AAAAAAGAAAAGTAGAATAATAAAA---AAAAAAAAAAAATAAAAATAAAAAAAAAAGA 384

Qy       1145 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1204

Db        383 AAAACATAAAAATAAAAATACAATAAAAAAAAAAAAAAAAAAAATAAAAAAAAAATTGACA 324

Qy       1205 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1264

Db        323 AACAAAAAAAAAAAAAAAGAAAAAAAATTAAACAAAAATTATTACGTAAAAAAAAAAAAG 264

Qy       1265 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1324

Db        263 ATTAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAGAAAAAAAAAAAATATA 204

Qy       1325 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1384

Db        203 AAAAAAAAAAAAAAAAAACAAAAATAGTAGAAAAAAAATGTACAAAAATAAAAACATATA 144

Qy       1385 TAAAATCATT 1394

Db        143 TAGTAGTAAT 134
```

Search completed: December 20, 2005, 23:02:14
Job time : 1620.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 15:08:19 ; Search time 408.691 Seconds
(without alignments)
1777.494 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161431 seqs, 245089505 residues

Total number of hits satisfying chosen parameters: 8322862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	77.6	5.2	169495	7	US-11-121-086-61	Sequence 61, Appl
c	2	74.4	5.0	3181	6	US-10-793-626-3866	Sequence 3866, Ap
c	3	71.4	4.8	3543	6	US-10-793-626-4149	Sequence 4149, Ap
c	4	70	4.7	171486	7	US-11-121-086-105	Sequence 105, App
	5	69.6	4.7	2655	6	US-10-793-626-1659	Sequence 1659, Ap
	6	69.6	4.7	3520	6	US-10-793-626-4160	Sequence 4160, Ap
	7	67.8	4.6	171486	7	US-11-121-086-105	Sequence 105, App
c	8	67.6	4.6	3602	6	US-10-793-626-3582	Sequence 3582, Ap
	9	67.6	4.6	15071	6	US-10-793-626-2963	Sequence 2963, Ap
c	10	67	4.5	173602	7	US-11-121-086-25	Sequence 25, Appl
	11	65.8	4.4	26772	6	US-10-995-561-13313	Sequence 13313, A
	12	65.8	4.4	54946	6	US-10-995-561-13479	Sequence 13479, A
	13	65.6	4.4	3840	6	US-10-793-626-3187	Sequence 3187, Ap
c	14	65.6	4.4	4965	6	US-10-485-517-43	Sequence 43, Appl
	15	64.4	4.3	6189	6	US-10-909-125-825	Sequence 825, App
	16	64	4.3	119036	6	US-10-995-561-13314	Sequence 13314, A
	17	63.6	4.3	201990	6	US-10-995-561-13303	Sequence 13303, A
c	18	63.2	4.3	1434	9	US-11-082-746-2	Sequence 2, Appli
	19	63	4.3	305312	6	US-10-995-561-13236	Sequence 13236, A
	20	62.8	4.2	627	6	US-10-793-626-693	Sequence 693, App
	21	62.8	4.2	627	6	US-10-793-626-1325	Sequence 1325, Ap
c	22	62.8	4.2	2999	6	US-10-793-626-4039	Sequence 4039, Ap
c	23	62.8	4.2	3691	6	US-10-793-626-3338	Sequence 3338, Ap
c	24	61.8	4.2	49979	6	US-10-995-561-13443	Sequence 13443, A
	25	61.6	4.2	148935	6	US-10-995-561-13308	Sequence 13308, A
	26	61.2	4.1	173602	7	US-11-121-086-25	Sequence 25, Appl
c	27	60.6	4.1	191091	7	US-11-121-086-60	Sequence 60, Appl
	28	60.2	4.1	2925	7	US-11-194-246-84	Sequence 84, Appl
c	29	59.6	4.0	415117	6	US-10-995-561-13274	Sequence 13274, A
c	30	59.4	4.0	3305	6	US-10-793-626-4002	Sequence 4002, Ap
	31	59.4	4.0	150468	7	US-11-112-908-56	Sequence 56, Appl
	32	59.4	4.0	193789	7	US-11-112-908-55	Sequence 55, Appl
c	33	59.2	4.0	2796	6	US-10-793-626-4335	Sequence 4335, Ap
c	34	58.8	4.0	1267	6	US-10-485-517-36	Sequence 36, Appl
c	35	58.8	4.0	118996	7	US-11-121-086-84	Sequence 84, Appl
	36	58.2	3.9	20317	6	US-10-995-561-13460	Sequence 13460, A
	37	57.6	3.9	98862	7	US-11-121-086-76	Sequence 76, Appl
	38	57.6	3.9	182190	7	US-11-121-086-102	Sequence 102, App
	39	57.6	3.9	394468	6	US-10-995-561-13473	Sequence 13473, A
	40	57.4	3.9	207908	7	US-11-112-908-21	Sequence 21, Appl
	41	57.4	3.9	212805	7	US-11-112-908-19	Sequence 19, Appl
c	42	57	3.8	151169	7	US-11-121-086-38	Sequence 38, Appl
	43	56.8	3.8	1752	6	US-10-793-626-1357	Sequence 1357, Ap
	44	56.8	3.8	3055	6	US-10-793-626-4075	Sequence 4075, Ap
	45	56.8	3.8	3543	6	US-10-793-626-3493	Sequence 3493, Ap

ALIGNMENTS

RESULT 1
US-11-121-086-61
; Sequence 61, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 169495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (70072)..(70171)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139457)..(157244)
; OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-61

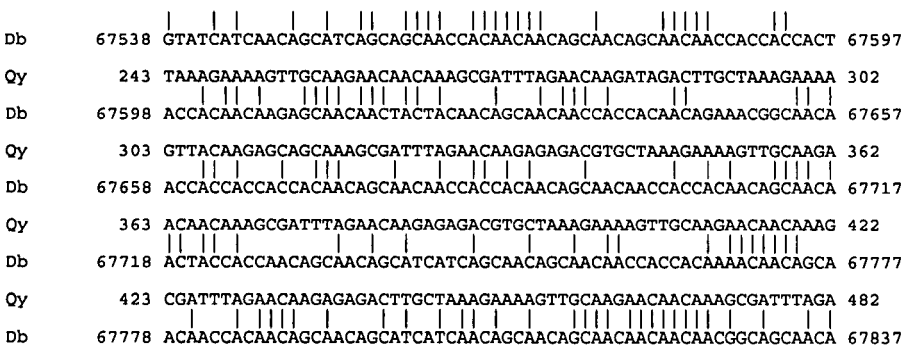
Query Match 5.2%; Score 77.6; DB 7; Length 169495;
Best Local Similarity 45.1%; Pred. No. 0.0055;
Matches 287; Conservative 0; Mismatches 349; Indels 0; Gaps 0;

```
Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              | | || |||     | | | ||  |   ||   |  ||  | ||   ||  |||||||
Db      67358 ACAGCAGCAACAACCACAAAAGCAGCAATAACTACCACAACGACAGCATCATCAACAACA 67417

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
                 ||  | ||||  ||   |       |  |  |   |||  |||| |||    |
Db      67418 GCAGCATCAACAACAGCATCATCAACAGCATCATCAACAACAGCAGCATCAACAACAGCG 67477

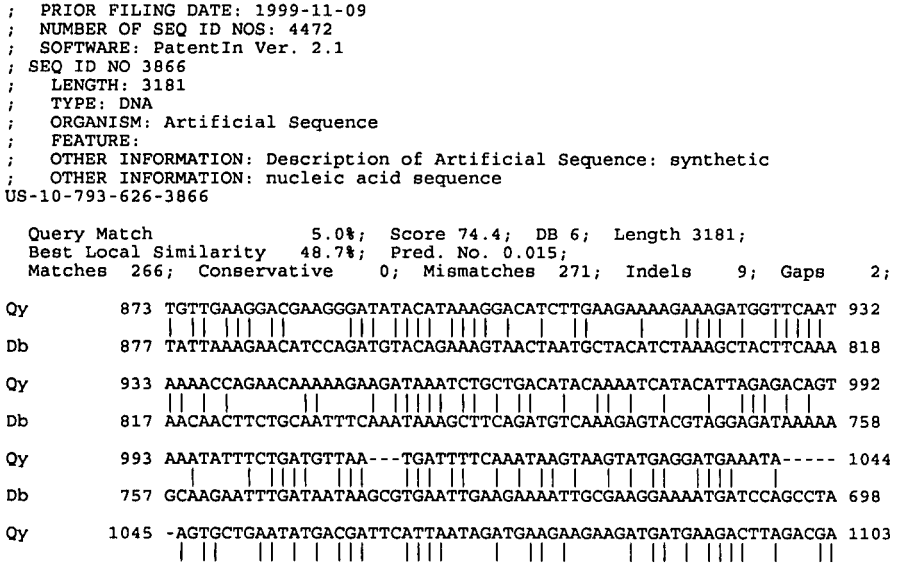
Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
              |||||  |             | || ||| ||||  |||| |      | | |  |||| |
Db      67478 TCAACAACAAGAGTAACAACCACAACAGTAGCAACCACAAAAGCAGCAACAACCACAACA 67537

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
```



RESULT 2
US-10-793-626-3866/c
; Sequence 3866, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3866
; LENGTH: 3181
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3866

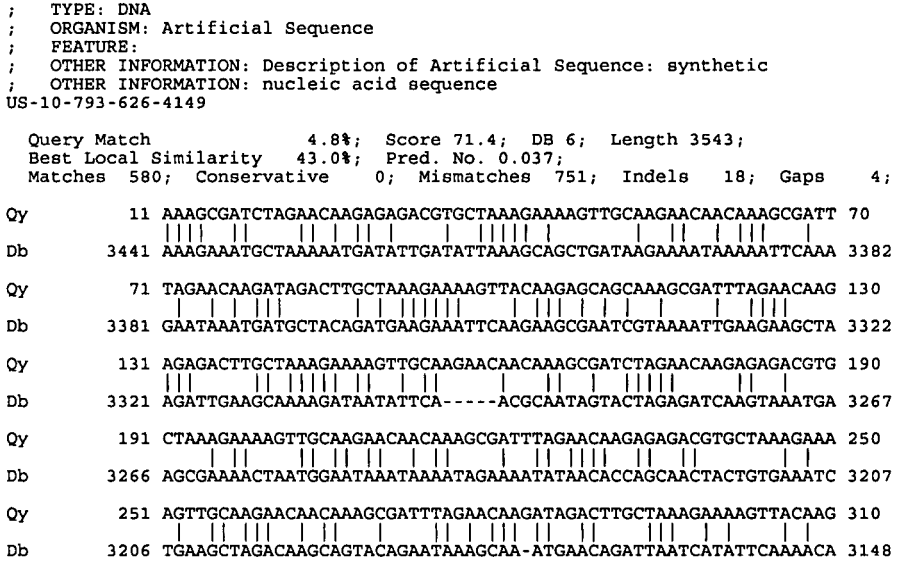
Query Match 5.0%; Score 74.4; DB 6; Length 3181;
Best Local Similarity 48.7%; Pred. No. 0.015;
Matches 266; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

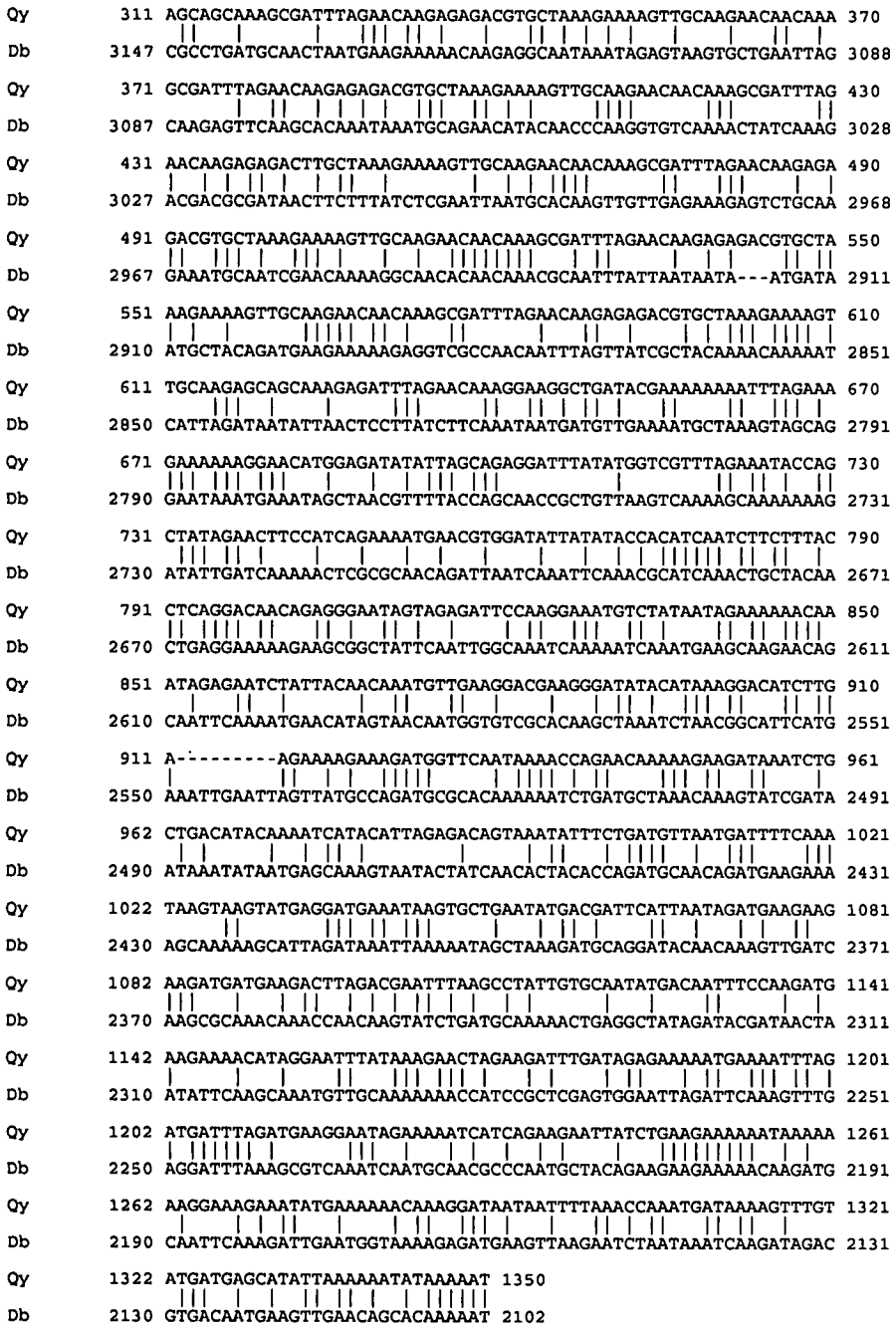




RESULT 3
US-10-793-626-4149/c
; Sequence 4149, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4149
; LENGTH: 3543
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4149

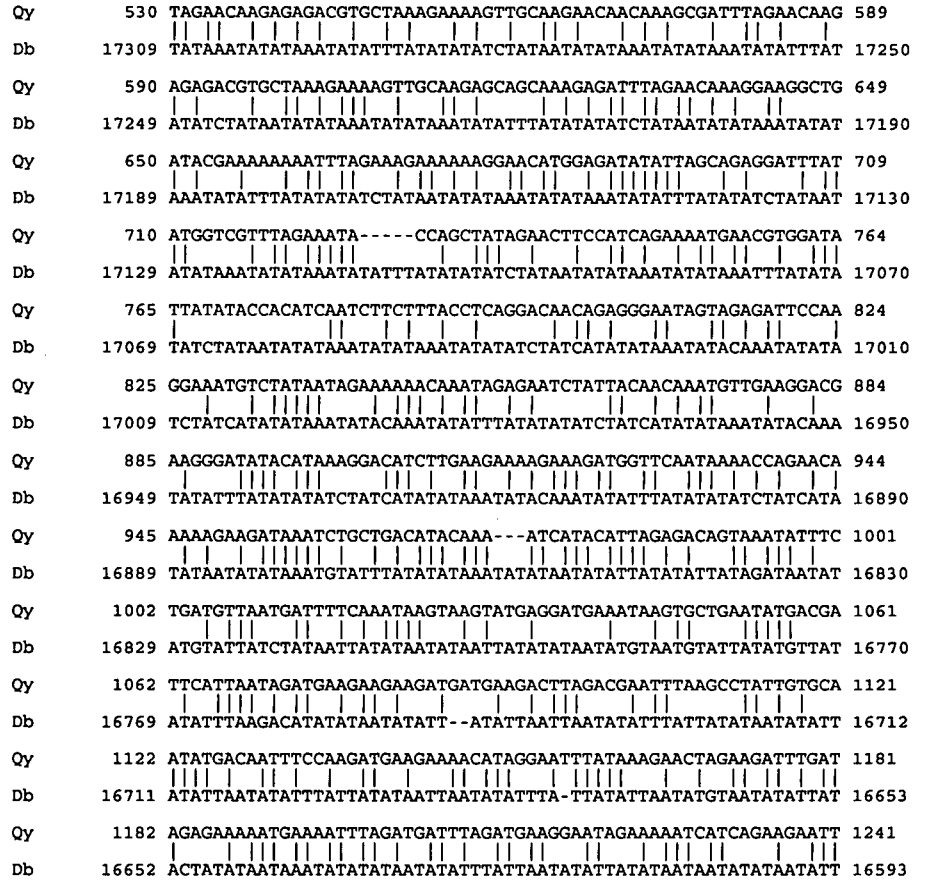
Query Match 4.8%; Score 71.4; DB 6; Length 3543;
Best Local Similarity 43.0%; Pred. No. 0.037;
Matches 580; Conservative 0; Mismatches 751; Indels 18; Gaps 4;

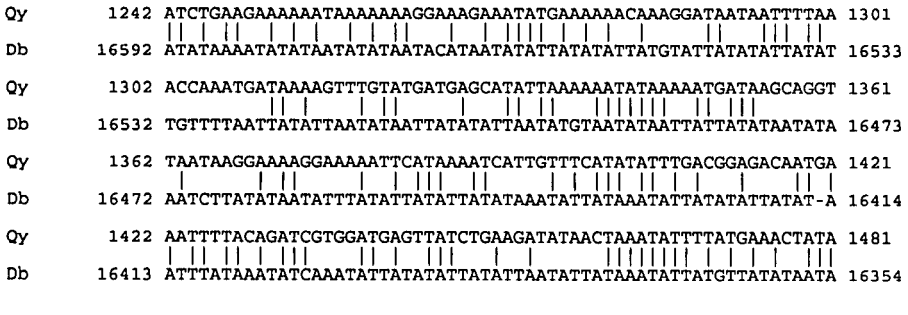




RESULT 4
US-11-121-086-105/c
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 171486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-105

Query Match 4.7%; Score 70; DB 7; Length 171486;
Best Local Similarity 45.6%; Pred. No. 0.058;
Matches 438; Conservative 0; Mismatches 510; Indels 12; Gaps 5;

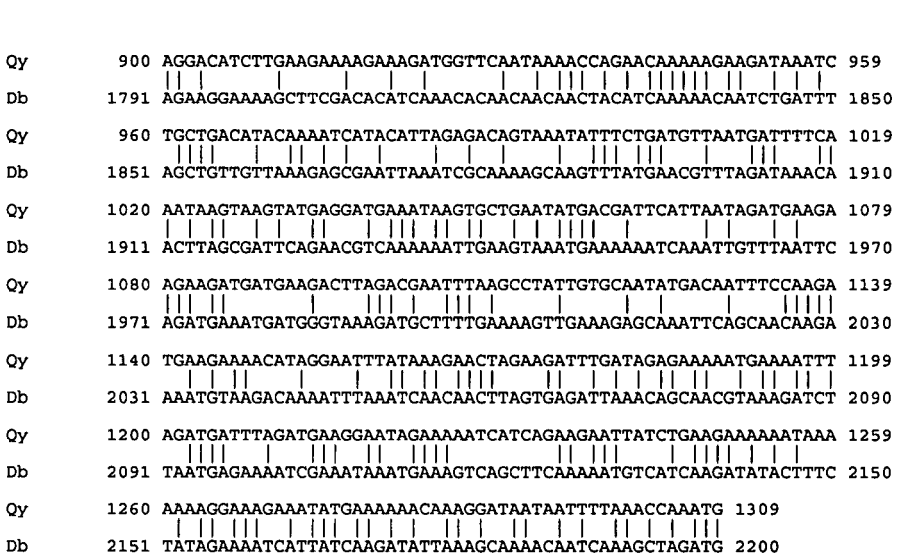




RESULT 5
US-10-793-626-1659
; Sequence 1659, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1659
; LENGTH: 2655
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1659

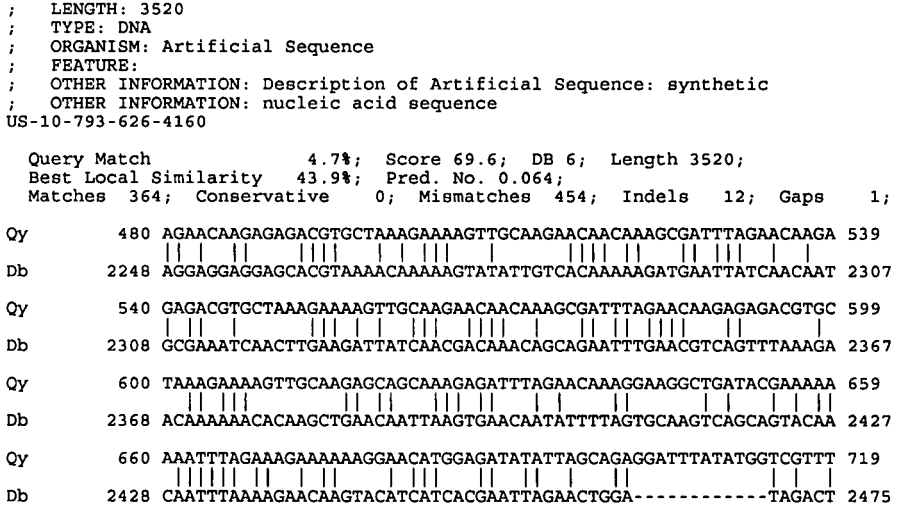
Query Match 4.7%; Score 69.6; DB 6; Length 2655;
Best Local Similarity 43.9%; Pred. No. 0.064;
Matches 364; Conservative 0; Mismatches 454; Indels 12; Gaps 1;





RESULT 6
US-10-793-626-4160
; Sequence 4160, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4160
; LENGTH: 3520
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4160

Query Match 4.7%; Score 69.6; DB 6; Length 3520;
Best Local Similarity 43.9%; Pred. No. 0.064;
Matches 364; Conservative 0; Mismatches 454; Indels 12; Gaps 1;



```
Db      2476 AAAAACACAAGAAGCACATCTTAAAAATGAACATGAAGAGTTTGAATTTGAAAAAAATGA 2535

Qy       780 ATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAAT 839

Db      2536 TGGATATCAAAGTGATAAAAGTAAAGAAACATTAAAAGAAAAACAAAATCATTTAATTGA 2595

Qy       840 AGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAA 899

Db      2596 GATACAACAACAATTGAAGCAACTAGAAAGTGATATTGAAAGATATACACAATTATCAAA 2655

Qy       900 AGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATC 959

Db      2656 AGAAGGAAAAGCTTCGACACATCAAACACAACAACAACTACATCAAAAACAATCTGATTT 2715

Qy       960 TGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCA 1019

Db      2716 AGCTGTTGTTAAAGAGCGAATTAAATCGCAAAAGCAAGTTTATGAACGTTTAGATAAACA 2775

Qy      1020 AATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGA 1079

Db      2776 ACTTAGCGATTCAGAACGTCAAAAAATTGAAGTAAATGAAAAAATCAAATTGTTTAATTC 2835

Qy      1080 AGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGA 1139

Db      2836 AGATGAAATGATGGGTAAAGATGCTTTTGAAAAGTTGAAAGAGCAAATTCAGCAACAAGA 2895

Qy      1140 TGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTT 1199

Db      2896 AAATGTAAGACAAAATTTAAATCAACAACTTAGTGAGATTAAACAGCAACGTAAAGATCT 2955

Qy      1200 AGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAA 1259

Db      2956 TAATGAGAAAATCGAAATAAATGAAAGTCAGCTTCAAAAATGTCATCAAGATATACTTTC 3015

Qy      1260 AAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG 1309

Db      3016 TATAGAAAATCATTATCAAGATATTAAAGCAAAACAATCAAAGCTAGATG 3065
```

RESULT 7
US-11-121-086-105
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 171486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-105

Query Match 4.6%; Score 67.8; DB 7; Length 171486;
Best Local Similarity 46.9%; Pred. No. 0.11;
Matches 277; Conservative 0; Mismatches 312; Indels 2; Gaps 2;

```
Qy       890 ATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAG 949

Db     16288 ATATTAATATAATATATAATATATAACATATAATATTTATATTAATATAATATATATTAT 16347

Qy       950 AAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTA 1009

Db     16348 ATGATATATTATATATAACATAATATTTATAATATTAATATAATATATAATATTTGATATTT 16407

Qy      1010 AT-GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTA 1068

Db     16408 ATAAATTATATAATATATAATATTTATAATATTTATATAATATAATATAAATATTATATA 16467

Qy      1069 ATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGAC 1128

Db     16468 AGATTTATATTATATAATAATTATATTACATATTAATATATAATTATATTAATATAATTA 16527

Qy      1129 AATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAAC-TAGAAGATTTGATAGAGAA 1187

Db     16528 AAACAATATAATATATAATACATAATATATAATATATTATGTATTATATATTATATATTT 16587

Qy      1188 AAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGA 1247

Db     16588 TATATAATATTATATATTATTATATAATATATTAATAAATATATTATATATATTTATTAT 16647

Qy      1248 AGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAA 1307

Db     16648 ATAGTATAATATATTACATATTAATATAATAAATATATTAATTATATAATAAATATATTA 16707

Qy      1308 TGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAA 1367

Db     16708 ATATAATATATTATATAATAAATATATTAATTAATATAATATATTATATATGTCTTAAAT 16767

Qy      1368 GGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTT 1427

Db     16768 ATATAACATATAATACATTACATATTATATATAATTATATTATATAATTATAGATAATAC 16827

Qy      1428 ACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db     16828 ATATATTATCTATAATATATAATATATTATATATTTATATATAAATACATT 16878
```

RESULT 8
US-10-793-626-3582/c
; Sequence 3582, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3582
; LENGTH: 3602
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3582

Query Match 4.6%; Score 67.6; DB 6; Length 3602;
Best Local Similarity 43.1%; Pred. No. 0.12;
Matches 438; Conservative 0; Mismatches 569; Indels 9; Gaps 2;

```
Qy       333 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACG 392

Db      2208 AGTAAATCCTGTTAAAGCAAGTTTAAATTATAGAAACGCTGATTATAATTTAAAACGTCA 2149

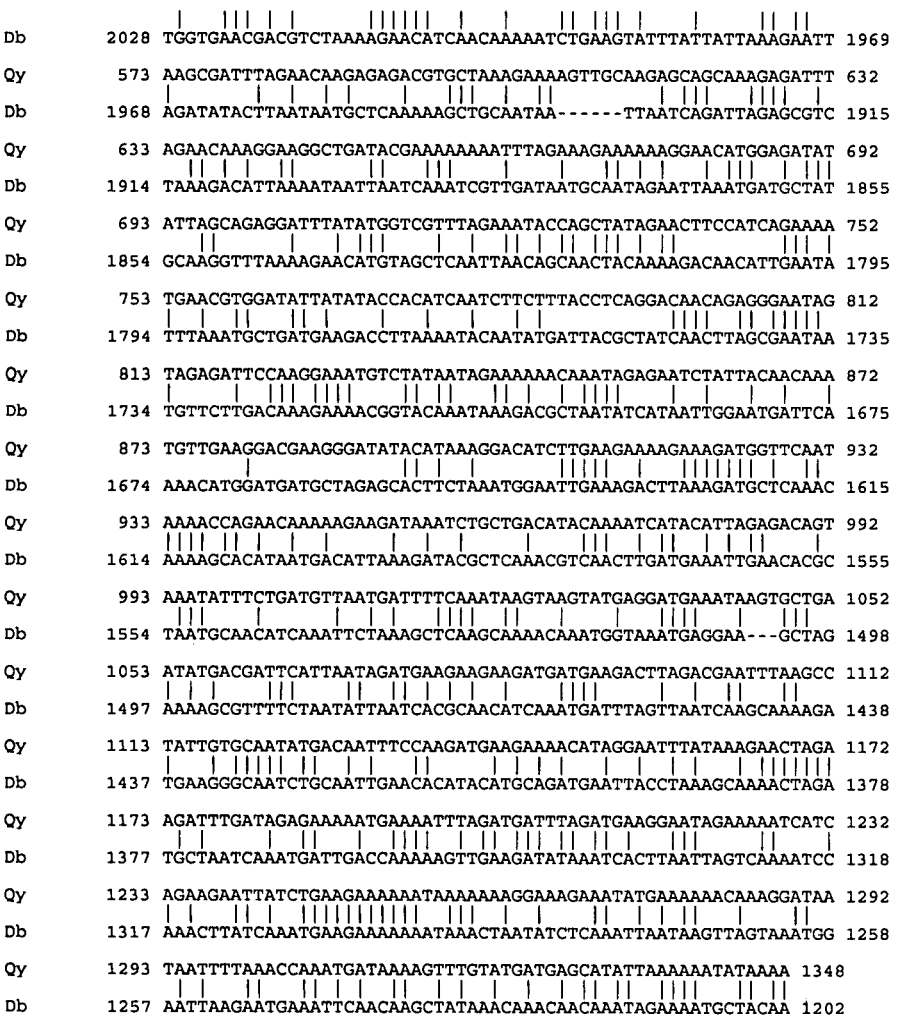
Qy       393 TGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGA 452

Db      2148 ATTTAACAAAGCTTTAAAAGAAGCTAAAGGCGTATTAAATAAAAATAGCGGTACAAATGT 2089

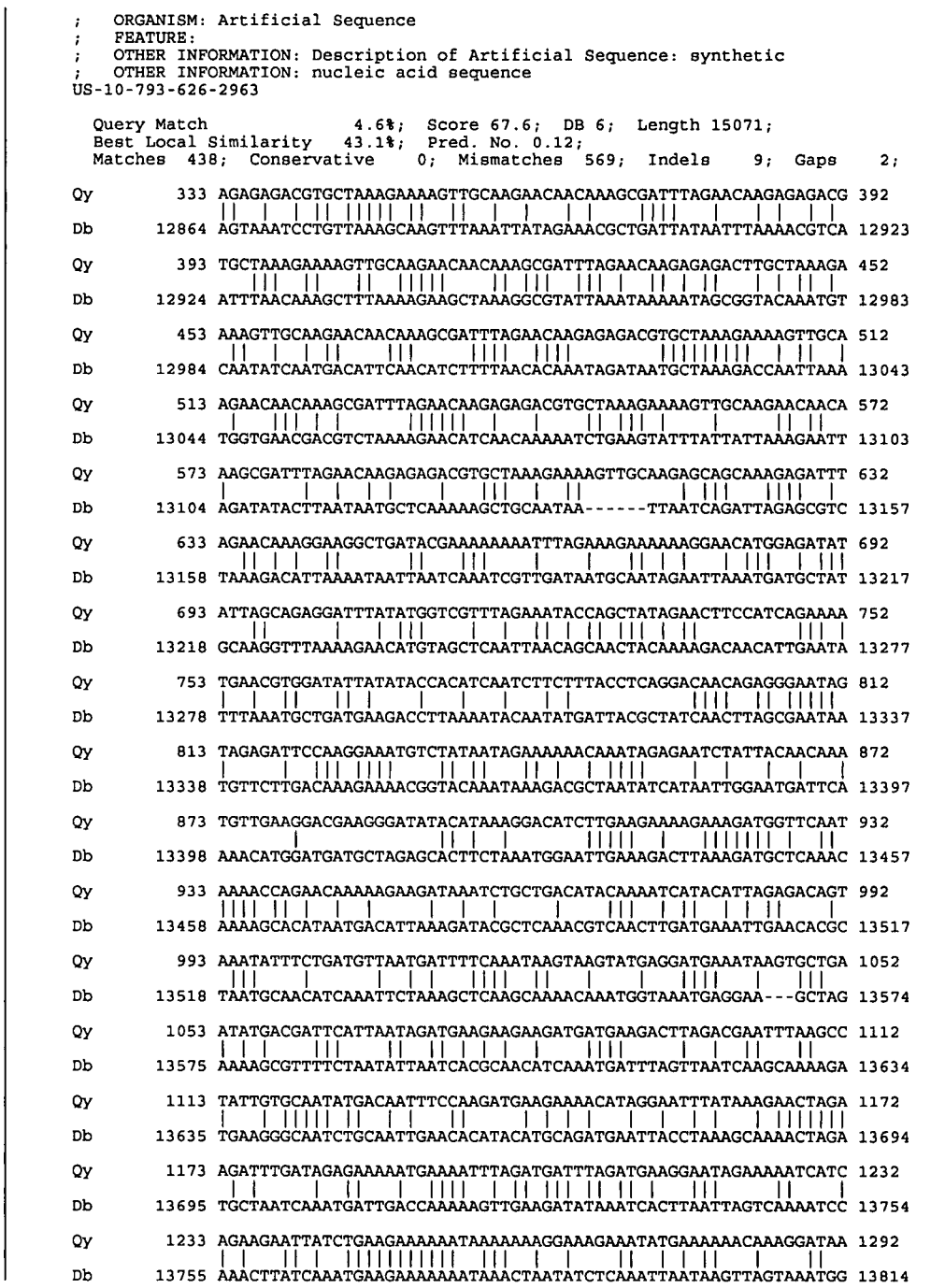
Qy       453 AAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 512

Db      2088 CAATATCAATGACATTCAACATCTTTTAACACAAATAGATAATGCTAAAGACCAATTAAA 2029

Qy       513 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 572
```



RESULT 9
US-10-793-626-2963
; Sequence 2963, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2963
; LENGTH: 15071
; TYPE: DNA



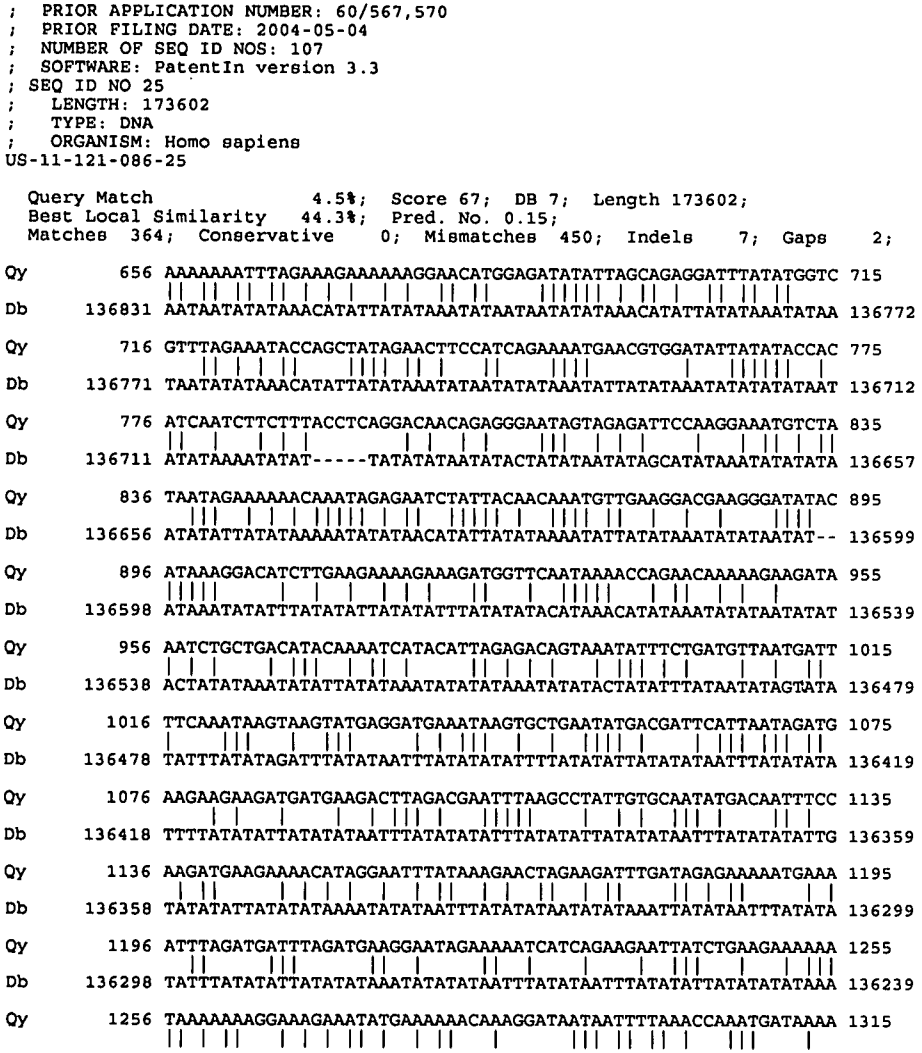
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2963

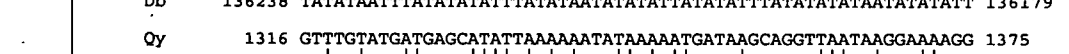
Query Match 4.6%; Score 67.6; DB 6; Length 15071;
Best Local Similarity 43.1%; Pred. No. 0.12;
Matches 438; Conservative 0; Mismatches 569; Indels 9; Gaps 2;



RESULT 10
US-11-121-086-25/c
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25

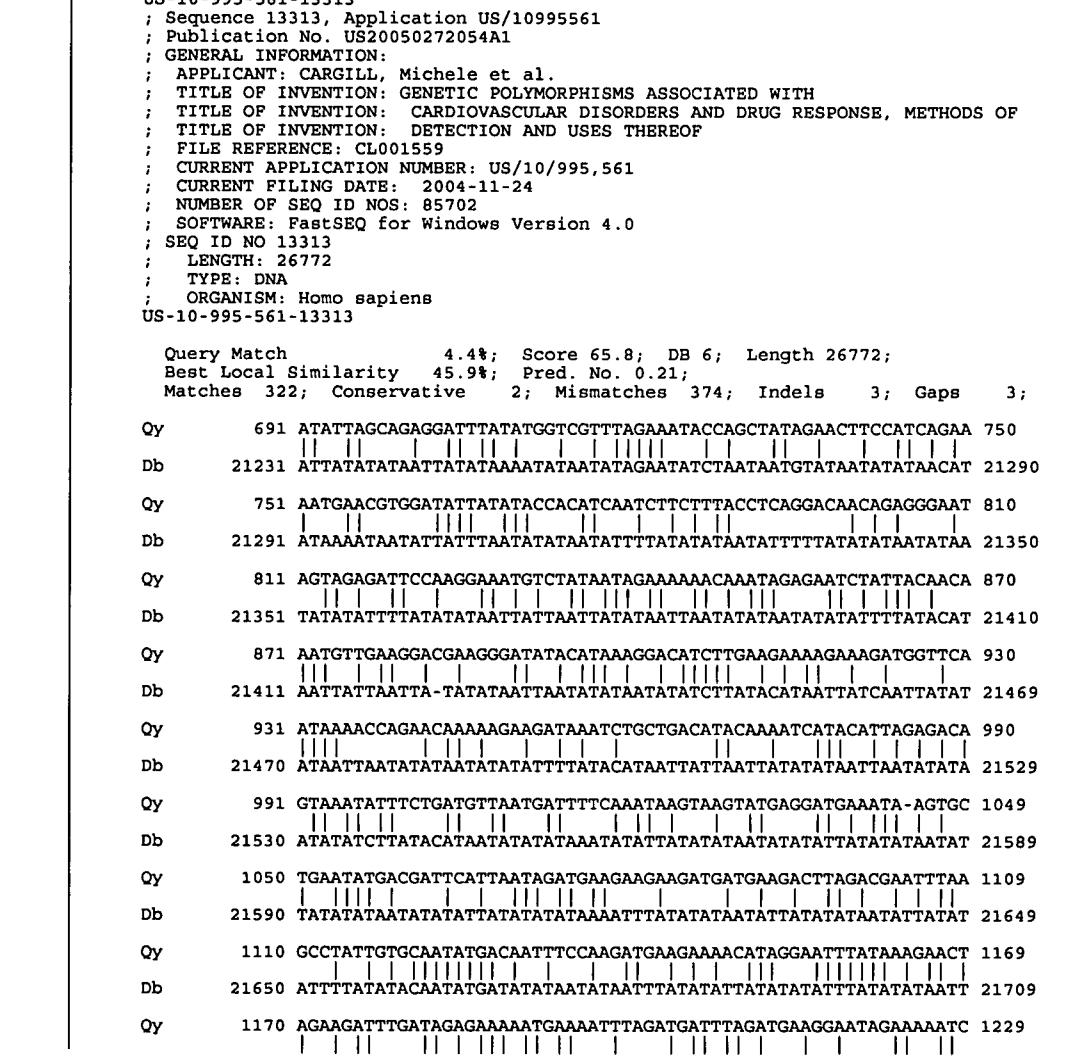
Query Match 4.5%; Score 67; DB 7; Length 173602;
Best Local Similarity 44.3%; Pred. No. 0.15;
Matches 364; Conservative 0; Mismatches 450; Indels 7; Gaps 2;





RESULT 11
US-10-995-561-13313
; Sequence 13313, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13313
; LENGTH: 26772
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13313

Query Match 4.4%; Score 65.8; DB 6; Length 26772;
Best Local Similarity 45.9%; Pred. No. 0.21;
Matches 322; Conservative 2; Mismatches 374; Indels 3; Gaps 3;



```
Db      21710 ATTATATAAATTATATAAATATAAATTATATATTTATATATAATTATTATATAAATCATT 21769

Qy       1230 ATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGA 1289

Db      21770 AT-ATAATTATTATAATTATAATATATAATATAATATAATATTATATATAMTATATAGTA 21828

Qy       1290 TAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAA 1349

Db      21829 TTCTATATAAATAATATAACATATATTTTATATAGAATATTATATATAATATAATATATA 21888

Qy       1350 TGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAAT 1390

Db      21889 TTTTATATAGAATATTATATATAATATAATATAYATTTTAT 21929


RESULT 12
US-10-995-561-13479
; Sequence 13479, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13479
;   LENGTH: 54946
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13479

  Query Match             4.4%;  Score 65.8;  DB 6;  Length 54946;
  Best Local Similarity   45.9%;  Pred. No. 0.21;
  Matches  322;  Conservative    2;  Mismatches  374;  Indels    3;  Gaps    3;

Qy        691 ATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAA 750

Db       2685 ATTATATATAATTATATAAAATATAATATAGAATATCTAATAATGTATAATATATAACAT 2744

Qy        751 AATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT 810

Db       2745 ATAAAATAATATTATTTAATATATAATATTTTATATATAATATTTTTATATATAATATAA 2804

Qy        811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACA 870

Db       2805 TATATATTTTATATATAATTATTAATTATATAATTAATATATAATATATTTTATACAT 2864

Qy        871 AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCA 930

Db       2865 AATTATTAATTA-TATATAATTAATATATAATATATCTTATACATAATTATCAATTATAT 2923

Qy        931 ATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACA 990

Db       2924 ATAATTAATATATAATATATATTTTATACATAATTATTAATTATATATAATTAATATATA 2983

Qy        991 GTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA-AGTGC 1049

Db       2984 ATATATCTTATACATAATATATATAAATATATTATATATAATATATTATATATAATAT 3043

Qy       1050 TGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAA 1109

Db       3044 TATATATAATATATATTATATATATAAAATTTATATATAATATTATATATAATATTATAT 3103

Qy       1110 GCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACT 1169

Db       3104 ATTTTATATACAATATGATATATAATATAATTTATATATTATATATATTTATATATAATT 3163

Qy       1170 AGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATC 1229

Db       3164 ATTATATAAATTATATAAATATAAATTATATATTTATATATAATTATTATATAAATCATT 3223

Qy       1230 ATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGA 1289

Db       3224 AT-ATAATTATTATAATTATAATATATAATATAATATAATATTATATATAMTATATAGTA 3282

Qy       1290 TAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAA 1349

Db       3283 TTCTATATAAATAATATAACATATATTTTATATAGAATATTATATATAATATAATATATA 3342

Qy       1350 TGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAAT 1390

Db       3343 TTTTATATAGAATATTATATATAATATAATATAYATTTTAT 3383


RESULT 13
US-10-793-626-3187
; Sequence 3187, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3187
;   LENGTH: 3840
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3187

  Query Match             4.4%;  Score 65.6;  DB 6;  Length 3840;
  Best Local Similarity   43.3%;  Pred. No. 0.22;
  Matches  424;  Conservative    0;  Mismatches  544;  Indels   12;  Gaps    2;

Qy        380 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 439

Db        137 AAGCACAAATAAATGCAGAACATACAACCCAAGGTGTCAAAACTATCAAAGACGACGCGA 196

Qy        440 GACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 499

Db        197 TAACTTCTTTATCTCGAATTAATGCACAAGTTGTTGAGAAAGAGTCTGCAAGAAATGCAA 256

Qy        500 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGT 559

Db        257 TCGAACAAAAGGCAACACAACAAACGCAATTTATTAATAATA---ATGATAATGCTACAG 313

Qy        560 TGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGC 619

Db        314 ATGAAGAAAAAGAGGTCGCCAACAATTTAGTTATCGCTACAAAACAAAAATCATTAGATA 373

Qy        620 AGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGG 679

Db        374 ATATTAACTCCTTATCTTCAAATAATGATGTTGAAAATGCTAAAGTAGCAGGAATAAATG 433

Qy        680 AACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAAC 739

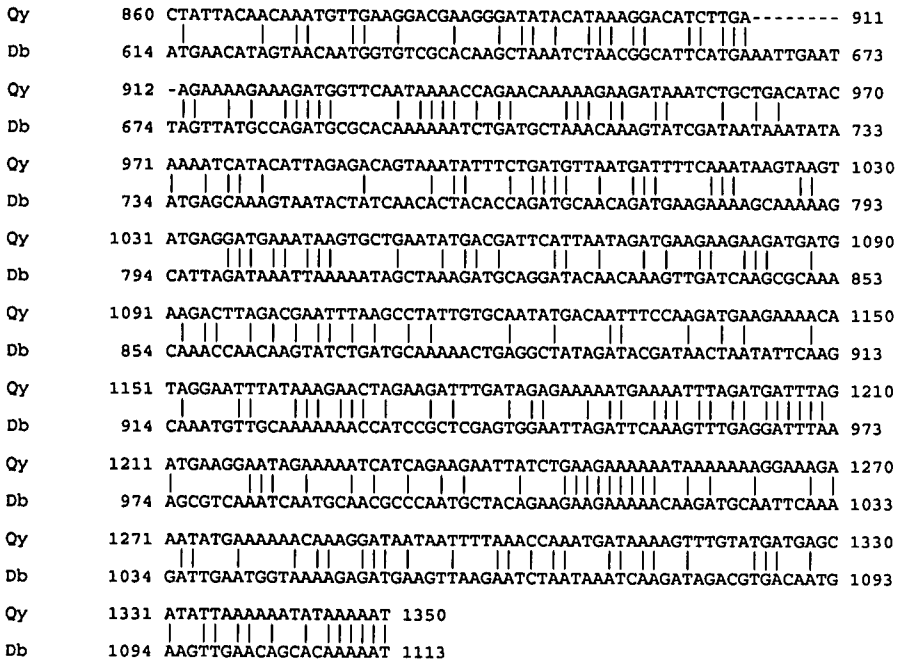
Db        434 AAATAGCTAACGTTTTACCAGCAACCGCTGTTAAGTCAAAAGCAAAAAAAGATATTGATC 493

Qy        740 TTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACA 799

Db        494 AAAAACTCGCGCAACAGATTAATCAAATTCAAACGCATCAAACTGCTACAACTGAGGAAA 553

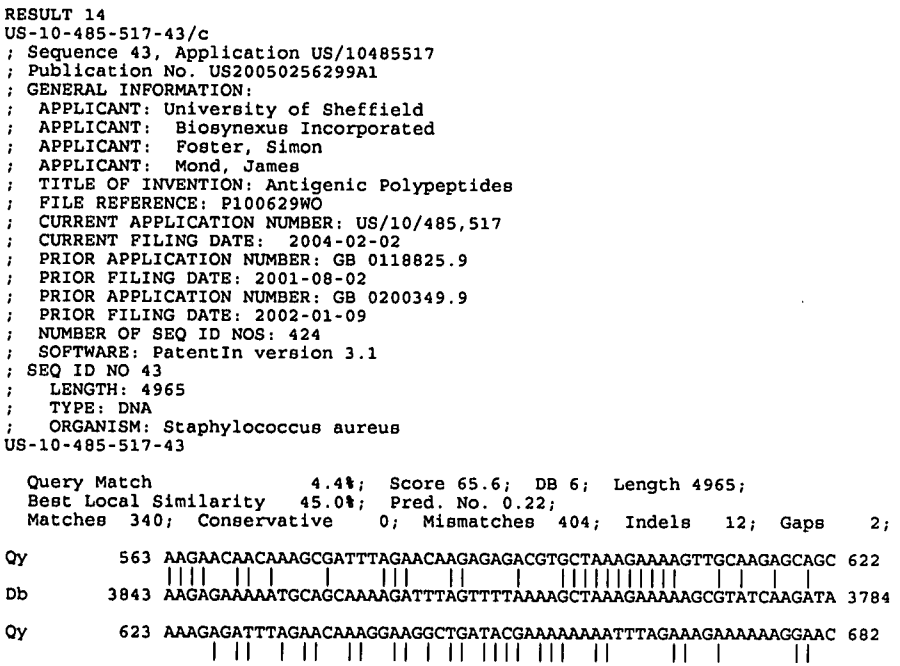
Qy        800 ACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAAT 859

Db        554 AAGAAGCGGCTATTCAATTGGCAAATCAAAAATCAAATGAAGCAAGAACAGCAATTCAAA 613
```



RESULT 14
US-10-485-517-43/c
; Sequence 43, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
; LENGTH: 4965
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-485-517-43

Query Match 4.4%; Score 65.6; DB 6; Length 4965;
Best Local Similarity 45.0%; Pred. No. 0.22;
Matches 340; Conservative 0; Mismatches 404; Indels 12; Gaps 2;



RESULT 15
US-10-909-125-825
; Sequence 825, Application US/10909125
; Publication No. US20050261218A1
; GENERAL INFORMATION:
; APPLICANT: Esau, Christine
; APPLICANT: Lollo, Bridget
; APPLICANT: Bennett, C. Frank
; APPLICANT: Freier, Susan M.
; APPLICANT: Griffey, Richard H.
; APPLICANT: Baker, Brenda F.
; APPLICANT: Vickers, Timothy
; APPLICANT: Marcusson, Eric G.
; APPLICANT: Koller, Erich
; APPLICANT: Swayze, Eric
; APPLICANT: Jain, Ravi
; APPLICANT: Bhat, Balkrishen
; APPLICANT: Peralta, Eigen
; TITLE OF INVENTION: Oligomeric Compounds And Compositions For Use In Modulation
; TITLE OF INVENTION: Of Small Non-Coding RNAs
; FILE REFERENCE: ISIS0080-100 (CORE0016US)
; CURRENT APPLICATION NUMBER: US/10/909,125
; CURRENT FILING DATE: 2004-07-30
; PRIOR APPLICATION NUMBER: US 60/492,056
; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: US 60/516,303

```
;  PRIOR FILING DATE: 2003-10-31
;  PRIOR APPLICATION NUMBER: US 60/531,596
;  PRIOR FILING DATE: 2003-12-19
;  PRIOR APPLICATION NUMBER: US 60/562,417
;  PRIOR FILING DATE: 2004-04-14
;  NUMBER OF SEQ ID NOS: 2184
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 825
;   LENGTH: 6189
;   TYPE: DNA
;   ORGANISM: H. sapiens
US-10-909-125-825

  Query Match             4.3%;  Score 64.4;  DB 6;  Length 6189;
  Best Local Similarity   42.4%;  Pred. No. 0.32;
  Matches  517;  Conservative    0;  Mismatches  696;  Indels    5;  Gaps    3;

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
              | | |||  ||    | | |     |    || |||  |||  ||||    | | |
Db       1894 ACATTTGAAAATTCAGATTGCACACTGAAAAAAATGGATAAAGAAGGTAAAACATTAAAA 1953

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
              ||| | || ||| | |  || | || ||   |||| | | | |   | |  |||
Db       1954 AAACATAAATTGAAGCATAAAGAGAGGGAAAAAGAAAAGCATAAAAAAGAAATTGAAGGT 2013

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
              | |||  || | |    |  ||    | | |     |   |  |  | ||
Db       2014 GAAAAGGAAAAATACAAAACTAAGGATAGTGCCAAAGAACTGCAGAGGAGTGTGGAATTT 2073

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
              |   |      ||   || | |      |   | ||  |     |||| |
Db       2074 GATAGAGAATTTTGGAAAGAGAATTTTTTTAAAAGTGATGAAACTGAAGATCTCTTTTTA 2133

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
              || | ||||| ||         |   ||||||      | ||    |||   | |  ||
Db       2134 AATATGGAACATGAATCCTTAACATTAGAAAAAAAATCAAAATTGGAAAAAAACATCAAA 2193

Qy        433 CAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 492
              | || | |    | || ||||||        | || | ||  | ||  || | ||  ||
Db       2194 GATGATAAATCAACCAAGGAAAAGCATGTGTCAAAAGAGAGGAACTTTAAAGAGGAACGA 2253

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                  |  |||||       ||  ||           ||| ||   | |   |  |||
Db       2254 GACAAGATTAAAAAGGAAAGCGAGAAATCTTTTAGGGAGGAAAAAATAAAAGATCTAAAA 2313

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 612
              ||| ||    || |   || |  |||  |||   |||       | ||    | |  | |
Db       2314 GAAGAGAGAGAAAACATACCCACAGATAAAGACTCAGAATTTACTTCTTTGGGTATG-AG 2372

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              |       ||   ||   |    |   |  |    |    || | | ||| | |||  |
Db       2373 TGCCATTGAGGAATCTATAGGGCTTCATTTAGTGGAAAAGGAAATAGACATTGAAAAACA 2432

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCT 732
              | ||| | | ||  ||  |  | | |  | |    | | | |     | | |   |  |
Db       2433 AGAAAAGCATATAAAGGAAAGTAAAGAAAAACCTGAGAAGCGATCTCAAATTAAAGAAAA 2492

Qy        733 ATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCT 792
              | |       |   | |||  ||      |   | | | | |     | | | |  | | |
Db       2493 GGACATTGAGAAGATGGAAAGAAAAACCTTTGAAAAAGAAAAGAAGATAAAACATGAGCA 2552

Qy        793 CAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAAT 852
              | | ||  | | |  | | |||| ||  ||  ||||| |  |  | | |||| |||
Db       2553 TAAGTCAGAAAAAGACAAATTAGATCTT---AGTGAATGTGTTGATAAAATAAAAGAAAA 2609

Qy        853 AGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
              || || ||||     ||         | |       | | |     || | | |   ||
Db       2610 GGACAAGCTATATTCGCATCACACAGAAAAATGCCATAAAGAAGGTGAGAAGAGCAAAAA 2669

Qy        913 GAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT-GACATACA 971
              |         ||    || | |   | || | | | |   | | || |    |
Db       2670 TACTGCTGCTATTAAAAAAACTGACGACAGAGAGAAAAGTAGAGAAAAGATGGATAGGAA 2729

Qy        972 AAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTA 1031
              | || | | |  | || | | |||         || |    ||     ||  ||   | |
Db       2730 ACATGACAAAGAAAAGCCTGAAAAAGAGAGGCATCTAGCAGAAAGCAAAGAAAAGCACTT 2789

Qy       1032 TGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGA 1091
                |||  ||| || |    |||  ||  ||       |  ||  |||| | |   |
Db       2790 GATGGAGAAAAAAAATAAACAATCAGATAATAGTGAATACAGTAAATCAGAAAAAGGCAA 2849

Qy       1092 AGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACAT 1151
              | |   |||  ||   | |    | |   || | || || |    |||| ||||||  ||
Db       2850 AAATAAAGAAAAAGACAGGGAGCTAGATAAAAAGGAAAAATCTAGAGATAAAGAAAGTAT 2909

Qy       1152 AGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGA 1211
              |   ||   |||       | |       |  ||| |||||   || | |||  ||     |
Db       2910 AAATATAACTAACTCCAAACACATACAGGAAGAAAAAAAATCAAGTATAGTAGACGGTAA 2969

Qy       1212 TGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAA 1271
              | ||| |  | |  ||  | |   |    |  ||| ||| || | |  ||     |||
Db       2970 TAAAGCACAACATGAAAAACCCTTATCCCTTAAAGAAAAAACAAAAGATGAACCTTTGAA 3029

Qy       1272 ATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCA 1331
              | | | |   ||| || ||  ||   |||   || |||| |  |   ||  | || |
Db       3030 AACTCCAGATGGAAAAGAAAAAGATAAAAAAGATAAAGATATAGATAGATACAAAGAACG 3089

Qy       1332 TATTAAAAAAATATAAAAA 1349
                 ||| | |  | ||
Db       3090 AGACAAACATAAAGATAA 3107


RESULT 16
US-10-995-561-13314
; Sequence 13314, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13314
;   LENGTH: 119036
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(119036)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13314

  Query Match             4.3%;  Score 64;  DB 6;  Length 119036;
  Best Local Similarity   45.5%;  Pred. No. 0.37;
  Matches  303;  Conservative    1;  Mismatches  356;  Indels    6;  Gaps    2;

Qy        814 AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
              | || ||  ||| |||  |       |  |||||    |  |   | ||||   | ||||
Db      31214 AAAGTTTAAAAGAAAAAATTATATTTTAGAAAAATGGCTGTATTTTTTATTTAGAAAAAT 31273

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
              |      | | |  | || ||| |    |  |  |   |||  |||||  ||  |||
Db      31274 GGCTGTATAAATAAGATTTTATATATATATAAAATATATATAAAATAAGAT-TTTATATA 31332

Qy        934 AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTA 993
              |   | || | | | || ||||   |    | ||| | ||         || |   ||
Db      31333 TATATAAAATATATATAAAATAAGATTTTATATATATATAAAATATATATAAAATAAGAT 31392

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
              |||  | |      |||   | | |||||||     |||    ||  |||  |    ||
Db      31393 TTTATATATATATAAAATATATATAAAATAAGATTTTATATATATAAAATATATATAA 31452
```

```
Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113
Db      31453 AATAAGATTTTATATATATATAAAATATATATAAAATAATATATAAAATAAGATTTCATA 31512

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAG-- 1171
Db      31513 TATATAAAATATATAAAATATATATAAAATATGATTTTTTATATATAAAAATATATATAA 31572

Qy       1172 ---AAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAAT 1228
Db      31573 AATAAGATTTTATATATAAAATATATAAAATATATATAAAATAAGATTTTATATATATAA 31632

Qy       1229 CATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGG 1288
Db      31633 AATATGATTTTATATATATAAAATATATATAAAATATATATAAAATAAGATTTTATATAT 31692

Qy       1289 ATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAA 1348
Db      31693 ATAAAATATATATAAAAATATATAAAATAAGATTTTATATATATATGAAATRTAAATAAG 31752

Qy       1349 ATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTG 1408
Db      31753 ATTTTATATATATATGAAATGTAAATAAGATTTTATATATTTGAAATATAAATAAAATAA 31812

Qy       1409 ACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATT 1468
Db      31813 GATTATATCTAAATATATATAAAATAAGATTTTACATATATGAAATATATATAAAATAAG 31872

Qy       1469 TTATGA 1474
Db      31873 GTTTTA 31878
```

RESULT 17
US-10-995-561-13303
; Sequence 13303, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13303
; LENGTH: 201990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(201990)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13303

Query Match 4.3%; Score 63.6; DB 6; Length 201990;
Best Local Similarity 45.1%; Pred. No. 0.42;
Matches 237; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

```
Qy        879 AGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACC 938
Db      78483 AGAAGCAGGAAATATAAAAAATTTATATAATATATATAAAATATATTATATATAAATATA 78542

Qy        939 AGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATAT 998
Db      78543 TAAATATATAAAAATAAATATATATTATATATATAAAAATATATAAATATATATAAAAAT 78602

Qy        999 TTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGA 1058
Db      78603 ATATAAATATATTATATATAATATATTAAATATATTATATATAATATATTTATATAATAT 78662
```

```
Qy       1059 CGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGT 1118
Db      78663 ATATAATATATTATATAAATATATAAATATATAAAAATATATAAAATATATAAATATATA 78722

Qy       1119 GCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTT 1178
Db      78723 ATATATATTATATATTAATATATTATATATTATATATAAATATATAAAAATATAATATAT 78782

Qy       1179 GATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGA 1238
Db      78783 ATAAAATATATATAAAAATATATGATATATAATATATATAAATATATGTATAAAATAT 78842

Qy       1239 ATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTT 1298
Db      78843 ATATTATATATTATATACTATATATTATATATAATATATAAAAAATATATATAAAATATA 78902

Qy       1299 TAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCA 1358
Db      78903 TATATAACTTCACGGAAACTCTGAAGGAGAGGTCAAATATAATATTATATATTTTATATA 78962

Qy       1359 GGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
Db      78963 TATAATTTATATATATATAAAACTATATATATATATTTTTGAGAGA 79008
```

RESULT 18
US-11-082-746-2/c
; Sequence 2, Application US/11082746
; Publication No. US20050244861A1
; GENERAL INFORMATION:
; APPLICANT: Sundstrom, Paula
; TITLE OF INVENTION: NUCLEIC ACIDS ESSENTIAL FOR EXPRESSION OF HYPHAL-SPECIFIC GENES
; TITLE OF INVENTION: AND METHODS FOR USING THE SAME
; FILE REFERENCE: DC-0300
; CURRENT APPLICATION NUMBER: US/11/082,746
; CURRENT FILING DATE: 2005-03-17
; PRIOR APPLICATION NUMBER: US 09/725,010
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: US 60/167,672
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Candida albicans
US-11-082-746-2

Query Match 4.3%; Score 63.2; DB 9; Length 1434;
Best Local Similarity 52.6%; Pred. No. 0.46;
Matches 160; Conservative 0; Mismatches 143; Indels 1; Gaps 1;

```
Qy       1107 TAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGA 1166
Db        305 TTAAACCACTGAATAAGCTTAGAACTAACTATTGTAAAAAAAAAGGGGTATTTATTAAAA 246

Qy       1167 ACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAA 1226
Db        245 GCTATCAACTATGAACCGAAAACAGAAAACAAAGCAAGTAGAAGAAGAAGAAAAAGATAA 186

Qy       1227 ATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAA 1286
Db        185 ACAATTTGAACAAAACAATTTTTAAAAACTAAAAGTACCTATCACCTTTAATGTAGTAAA 126

Qy       1287 GGATAATAATTTTAAACCAAATGATAAAAGTTTGT-ATGATGAGCATATTAAAAAATATA 1345
Db        125 CAAAACAAATATTCAATAAAATGACTCTTGTCAATAATATTGAAAACCCAAAATTGAAAA 66

Qy       1346 AAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATAT 1405
Db          65 ACAATCAAAATTAGAAAAAAAAGAAATCCCAAAAAAAGAACAAAACTTAGTATCAGTTAT 6

Qy       1406 TTGA 1409
```

```
Db          5 TAGA 2


RESULT 19
US-10-995-561-13236
; Sequence 13236, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13236
;   LENGTH: 305312
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(305312)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13236

  Query Match             4.3%;  Score 63;  DB 6;  Length 305312;
  Best Local Similarity   51.6%;  Pred. No. 0.51;
  Matches  144;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy       1155 AATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGA 1214
Db     140661 AATAAAATAAATAACATAACATAACATAACATAACATAACATAACATAACATATAAAATA 140720

Qy       1215 AGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATA 1274
Db     140721 AAATAAAATATAAAATAAAATAAAATAAAATAAAATAAAAATAAAATAAAATATAAAATA 140780

Qy       1275 TGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATAT 1334
Db     140781 AAATAAAATAAAATAAAATAAATAAAATAAAATAAAATAAAATATAAAATAAAATAAAAT 140840

Qy       1335 TAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATT 1394
Db     140841 AAAATAAAATAAAAATAAAATAAAATAAAATAAAATAAAATATAAAATAAAATAAATAA 140900

Qy       1395 GTTTCATATATTTGACGGAGACAATGAAATTTTACAGAT 1433
Db     140901 ATAAAATAAAATAAAATAAAATAAAATAAAAAAATAAAT 140939


RESULT 20
US-10-793-626-693
; Sequence 693, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 693
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-693

  Query Match             4.2%;  Score 62.8;  DB 6;  Length 627;
  Best Local Similarity   48.6%;  Pred. No. 0.52;
  Matches  203;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy       1061 ATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGC 1120
Db        134 ATCAAGTAAGTAAAAAAATAAACGAATTAGAAGAGAAAAAGCAAAAATTATTTAAAAAGG 193

Qy       1121 AATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGA 1180
Db        194 TAAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAA 253

Qy       1181 TAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAAT 1240
Db        254 AACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCAT 313

Qy       1241 TATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA 1300
Db        314 TTAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTG 373

Qy       1301 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGG 1360
Db        374 AACAACTTGATAGTGCTAT---TAAAGAAAAATATAAATCACATGATGCTTATGCAAAGG 430

Qy       1361 TTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATG 1420
Db        431 CTTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATG 490

Qy       1421 AAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478
Db        491 CAACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 548


RESULT 21
US-10-793-626-1325
; Sequence 1325, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1325
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1325

  Query Match             4.2%;  Score 62.8;  DB 6;  Length 627;
  Best Local Similarity   48.6%;  Pred. No. 0.52;
  Matches  203;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy       1061 ATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGC 1120
Db        134 ATCAAGTAAGTAAAAAAATAAACGAATTAGAAGAGAAAAAGCAAAAATTATTTAAAAAGG 193

Qy       1121 AATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGA 1180
Db        194 TAAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAA 253

Qy       1181 TAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAAT 1240
```

```
Db        254 AACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCAT 313

Qy       1241 TATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA 1300

Db        314 TTAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTG 373

Qy       1301 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGG 1360

Db        374 AACAACTTGATAGTGCTAT---TAAAGAAAAATATAAATCACATGATGCTTATGCAAAGG 430

Qy       1361 TTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATG 1420

Db        431 CTTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATG 490

Qy       1421 AAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db        491 CAACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 548


RESULT 22
US-10-793-626-4039/c
; Sequence 4039, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4039
;   LENGTH: 2999
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4039

  Query Match             4.2%;  Score 62.8;  DB 6;  Length 2999;
  Best Local Similarity   48.6%;  Pred. No. 0.53;
  Matches  203;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy       1061 ATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGC 1120

Db       1686 ATCAAGTAAGTAAAAAAATAAACGAATTAGAAGAGAAAAAGCAAAAATTATTTAAAAAGG 1627

Qy       1121 AATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGA 1180

Db       1626 TAAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAA 1567

Qy       1181 TAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAAT 1240

Db       1566 AACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCAT 1507

Qy       1241 TATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA 1300

Db       1506 TTAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTG 1447

Qy       1301 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGG 1360

Db       1446 AACAACTTGATAGTGCTAT---TAAAGAAAAATATAAATCACATGATGCTTATGCAAAGG 1390

Qy       1361 TTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATG 1420

Db       1389 CTTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATG 1330

Qy       1421 AAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db       1329 CAACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 1272


RESULT 23
US-10-793-626-3338/c
; Sequence 3338, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3338
;   LENGTH: 3691
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3338

  Query Match             4.2%;  Score 62.8;  DB 6;  Length 3691;
  Best Local Similarity   48.6%;  Pred. No. 0.53;
  Matches  203;  Conservative    0;  Mismatches  212;  Indels    3;  Gaps    1;

Qy       1061 ATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGC 1120

Db       2726 ATCAAGTAAGTAAAAAAATAAACGAATTAGAAGAGAAAAAGCAAAAATTATTTAAAAAGG 2667

Qy       1121 AATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGA 1180

Db       2666 TAAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAA 2607

Qy       1181 TAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAAT 1240

Db       2606 AACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCAT 2547

Qy       1241 TATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTA 1300

Db       2546 TTAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTG 2487

Qy       1301 AACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGG 1360

Db       2486 AACAACTTGATAGTGCTAT---TAAAGAAAAATATAAATCACATGATGCTTATGCAAAGG 2430

Qy       1361 TTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATG 1420

Db       2429 CTTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATG 2370

Qy       1421 AAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db       2369 CAACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 2312


RESULT 24
US-10-995-561-13443/c
; Sequence 13443, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13443
```

```
;   LENGTH: 49979
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(49979)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13443

  Query Match             4.2%;  Score 61.8;  DB 6;  Length 49979;
  Best Local Similarity   48.4%;  Pred. No. 0.73;
  Matches  220;  Conservative    3;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
              | | || ||| || |  |||  | | | || || | | | || || ||   | |  ||||
Db      29303 GGCCTAGAAACTCTTTTATTTTAAAAAATATATGTATATTATATTTATATATATATAATA 29244

Qy       1024 AGTAAGTATGAGGATGAAATAAGT-GCTGAATATGACGATTCATTAATAGATGAAGAAGA 1082
                |||  |||   ||  |||| ||   | ||||| | |  | | ||  | ||  | |
Db      29243 TATAAAAATGTATATATAATATGTATATAAATATAATGTATAAATATAATATTTATATAT 29184

Qy       1083 AGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGA 1142
              |  |    |    ||     | | ||    || | |  | |||    ||      | || |
Db      29183 ATTTATAAATATATTTATATATTATATTTATAATATATATTATATATATAATATATATAA 29124

Qy       1143 AGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGA 1202
                || | ||    | ||||| | ::     | | |  ||| |  | |    | |  || |
Db      29123 TAAATATATGTACAATTATATATWWTATATATAATATATATAATATATATTATAGATATA 29064

Qy       1203 TGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAA 1262
              | |  || ||  |   || | | || ||      | |      |  | | |  || ||
Db      29063 TTAAATATATATATTTTTATATATATTATATATTTATTATATATATTATATATAAATAAT 29004

Qy       1263 AGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTA 1322
              |  ||   |||||| |  |    |   ||||||    || |  | ||   |: | | |  |
Db      29003 ATAAATATAATATGTATTATATATAAATAATATAAATATAATATATATAAMCAATATAAA 28944

Qy       1323 TGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATT 1382
              |      |||| |||| |||||||  |   ||||     |  ||||  || |  || |
Db      28943 TATATTACATA-TAAATAATATAATATATATAATATATTATATAAATAATATAAATATAA 28885

Qy       1383 CATAAAATCATTGTTTCATATATTTGACGGAGACA 1417
                ||| | |   | | | |||||  | |   ||| |
Db      28884 TATATATTACATATGTAATATAAATAATATAGATA 28850


RESULT 25
US-10-995-561-13308
; Sequence 13308, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13308
;   LENGTH: 148935
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13308

  Query Match             4.2%;  Score 61.6;  DB 6;  Length 148935;
  Best Local Similarity   43.3%;  Pred. No. 0.78;
  Matches  257;  Conservative    6;  Mismatches  331;  Indels    0;  Gaps    0;

Qy         27 AGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACT 86
              | |||| ||| |||   | |      |||| ||        |  |  |   | | || |
Db      22401 ACAGAGGCGTCCTAGTCACACTGCTAAAGATCACAGTCTTAAAATCCATGCACAGAGGCR 22460

Qy         87 TGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGA 146
              | :||   | |       |||| || ||      |  |  |   | |||| | | ||    |
Db      22461 TCYTAGTCACACTGCTRAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTGGTCA 22520

Qy        147 AAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCA 206
               |    |||| || ||    |  |  |  | |||| ||| ||:   | |      |
Db      22521 CACTGCTAAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTRGTCACACTGCTGA 22580

Qy        207 AGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 266
              ||| ||  :     |  |  |   | |||| |:: |||   | |      |||| || |:
Db      22581 AGATCACARTCTTAAAATCCATGCACAGAGGCRYCCTAGTCACACTGCTGAAGATCACCV 22640

Qy        267 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 326
                 |  |  |  | | || |   |||   | |      |||| || ||      |  |
Db      22641 TCTTAAAATCCATGCACAGAGGCACCCTAGTCACACTGCTGAAGATCACCATCTTAAAAT 22700

Qy        327 AGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGA 386
               |  | |||| ||| |||   | |      |||| || ||      |  |     | |
Db      22701 CCATGCACAGAGGCGTCCTAGTCACACTGCTGAAGATCACCATCTTAAAATCCGTGCACA 22760

Qy        387 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGC 446
              ||| ||| |||   | |       |||| ||         |  |  |   | |||| |   |
Db      22761 GAGGCGTCCTAGTCACACTGCTGAAGATCACAGTCTTAAAATCCATGCACAGAGGCACCC 22820

Qy        447 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAA 506
              ||   | |       |||| || ||      |  |  |   | |||| ||| ||     | |
Db      22821 TAGTCACACTGCTGAAGATCACCATCTTAAAATCCATGCACAGAGGCGTCCTGGTCACAC 22880

Qy        507 GTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 566
                 |||| ||       |  |  |   | |||| |   |||   | |      ||||
Db      22881 TGCTGAAGATCACAGTCTTAAAATCCATGCACAGAGGCACCCTAGTCACACTGCTGAAGA 22940

Qy        567 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCA 620
              || ||      |  |  |   | |||| ||| |||   | |      |||| ||
Db      22941 TCACCATCTTAAAATCCATGCACAGAGGCGTCCTAGTCACACTGCTAAAGATCA 22994


RESULT 26
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
;   LENGTH: 173602
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-25

  Query Match             4.1%;  Score 61.2;  DB 7;  Length 173602;
  Best Local Similarity   44.7%;  Pred. No. 0.88;
  Matches  370;  Conservative    0;  Mismatches  448;  Indels   10;  Gaps    3;

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              | | |||  |   | | | |  |  | ||  | ||| ||||        || |||||  |
Db     136193 ATATAAATATATAATATATATTATATAAATATATATAAATTATATATTTATATATATAAT 136252

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                | |   |||  |  |||| |  ||      | || ||  |  |  |||| |  ||
Db     136253 ATATAAATTATATAAATTATATATATTTATATATAATATATAAATATATATAAATTATAT 136312

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCT 834
               ||  || |  | ||  | |     | |      |||  | | || | |   | ||   |
```

```
Db     136313 AATTTATATATTATATATAAATTATATATTTTATATATAATATATACAATATATATAAAT 136372

Qy        835 ATAATAGAAAAAACAAATAGAGA--------ATCTATTACAACAAATGTTGAAGGACGAA 886
              ||| || | | ||||| | |          || || || |  |||| |  |    |  |
Db     136373 TATATATAATATATAAATATATATAAATTATATATAATATATAAAATATATATAAATTAT 136432

Qy        887 GGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAA 946
              ||| |||||  | || |  | | | ||        |  |  | || | | | | |
Db     136433 ATATAATATATAAAATATATATAAATTATATAAATCTATATAAATATATACTATATTATA 136492

Qy        947 AAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATG 1006
              || |   || || |    | |||   |   |   | | |||| |||||   ||
Db     136493 AATATAGTATATATTTATATATATTTATATAATATATTTATATAGTATATATTATATATT 136552

Qy       1007 TTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCAT 1066
              |  ||| || |  |   |  ||    |  ||  ||  |    | |||| |    |
Db     136553 TATATGTTTATGTATATATAAATATATAATATATAAATATATTTATATATTATATATTTA 136612

Qy       1067 TAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATG 1126
              ||  | ||     | || || |  | | ||  |  | | |   |||    |||||
Db     136613 TATAATATTTTATATAATATGTTATATATTTTTATATAATATATTATATATATTTATATG 136672

Qy       1127 ACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGA-ACTAGAAGATTTGATAGAG 1185
              |  ||  |  |  | | | | |||  |  || || | |  || |    || ||| |
Db     136673 CTATATTATATAGTATATTATATATAATATATTTTATATATTATATATATATTTATATAA 136732

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
              |      | | | ||  |   | | |||      |||  |  ||  | | |  |  | | |
Db     136733 TATTTATATATTATATTTATATAATATGTTTATATATTATTATATTTATATAATATGTTT 136792

Qy       1246 GAAGA-AAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACC 1304
              | | |  |||   | | | | |    |||  |  |  | |     ||||| |
Db     136793 ATATATTATTATATTTATATAATATGTTTATATATTATTATATTTATATAATATGTTTAT 136852

Qy       1305 AAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAA 1364
              ||   ||| ||  |  |   || |    ||| | | |||| | || |||      |  |
Db     136853 AATATATATAATATATTTATATAATATGTTTATATATTATATATATAATATATTTATATA 136912

Qy       1365 TAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAAT 1424
              | |   |   |  ||  ||| |||   | | |    |||  |    |    || | ||
Db     136913 ATATGTTTATATATTATATATATAATATATTTATATATTATTATATATATTTTATAATAT 136972

Qy       1425 TTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTAT 1472
              | | | ||     || |  ||| |        | || |||| ||||
Db     136973 ATAATATATATAATATAATATATATATTATATATATTATATATATTAT 137020


RESULT 27
US-11-121-086-60/c
; Sequence 60, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
;   LENGTH: 191091
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-60

  Query Match             4.1%;  Score 60.6;  DB 7;  Length 191091;
  Best Local Similarity   49.3%;  Pred. No. 1.1;
  Matches  185;  Conservative    0;  Mismatches  189;  Indels    1;  Gaps    1;

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              |||  | |  || |      |||||||  |  | ||  | | |  ||   | |  |   |
Db      40243 GAAAGAAAAGGAAGAAGGGAAGAAAAGGAGAGAAAAAGAGAGAAGGAAAGAAAGGAGAGG 40184

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
              | | |  |||||||||     ||| | |  || | ||  |||| | |  |
Db      40183 GAAAGAGGGAAAGAAAAGAAAGAAG-AGAGAAAGGAGAGAAAGAAAAGAAAGAAGAGAGA 40125

Qy        448 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 507
              || ||||    | ||  | |  ||| ||  |    | |||  |        ||| ||||
Db      40124 AAGGAAATAAAGAAAAGATAGAAAAAAGAAGGAAGGAAGGAGGAAAAAAAGGAAGGAAAG 40065

Qy        508 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 567
              ||||   | | || ||  | |  |||| |||      ||  || |  | |||||
Db      40064 AGAGAAGAGAGAAAGAGAGAAAGAAAGAAAGAAAGAGAAAGAAAGAAAAGAAAGAA 40005

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
              || ||||     ||||  | ||| | |   | | | ||      | || |  || ||
Db      40004 AAAGAAAGAAGGAAAGAAAGAAAGAAAAGAAAAAGAAAGAAGGAAAGAAAGAAGGA 39945

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
              |  ||||  || ||| |      | |||| ||        ||  || |||| || |  ||
Db      39944 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAGGAAAGAAAGAAAGAAAGAAAGAAGA 39885

Qy        688 GATATATTAGCAGAG 702
              | | | |  ||||
Db      39884 AAGAAAGAAAGAGAG 39870


RESULT 28
US-11-194-246-84
; Sequence 84, Application US/11194246
; Publication No. US20050272089A1
; GENERAL INFORMATION:
;  APPLICANT: Mott, John
;  APPLICANT:  Trepod, Catherine
;  APPLICANT:  Arvidson, Staffan
;  TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND ME
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: 00592.US1 (M&R 268.05920101)
;  CURRENT APPLICATION NUMBER: US/11/194,246
;  CURRENT FILING DATE:  2005-08-01
;  PRIOR APPLICATION NUMBER: US/10/274,586
;  PRIOR FILING DATE: 2002-10-21
;  PRIOR APPLICATION NUMBER: US 60/345,438
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 621
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84
;   LENGTH: 2925
;   TYPE: DNA
;   ORGANISM: ARTIFICIAL
;   FEATURE:
;   OTHER INFORMATION: Sequence of the IF-2 coding sequence and flanking regions.
US-11-194-246-84

  Query Match             4.1%;  Score 60.2;  DB 7;  Length 2925;
  Best Local Similarity   44.1%;  Pred. No. 1.2;
  Matches  251;  Conservative    0;  Mismatches  318;  Indels    0;  Gaps    0;

Qy        812 GTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAA 871
              ||| | |   ||  | ||   ||   || |   ||  | |   || |     | | | ||
Db       1201 GTAAAAAACGCACAGTAAAAACTGATATTGCTCAACAAGAAGAAGCAAAATTAAAAGCAC 1260

Qy        872 ATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAA 931
              |    ||||    |  | | | |  |   | |||    | || | |  || | | |
Db       1261 AGCAAGAAGCGGAAGCGAAAAAAATTGCTGAACAAAAAGCAGCAGAAGAAAAAGCTCGTT 1320

Qy        932 TAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAG 991
              || || | ||| ||  | ||| | || ||||| |   |  |||   | |  ||||  |
Db       1321 TAGAAGCTGAAAAAGCAAAAGCTGAAACTGCTAAGCCAGTAAAAAGTGCGGTAGATTCTA 1380

Qy        992 TAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTG 1051
              |  ||  |||| ||    |||     | | | ||||    |  || | |  |  |  |
```

```
Db       1381 AAGCTAAATCTGTTGAGTCTGAAAAAGAGAAACGTAAAGCAGGAGAAGCAGAACTTCGTC 1440

Qy       1052 AATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGC 1111

Db       1441 GTAAAGCTGAAGAATTAGCTCGTCAAAAAGCAGAAGAACAAGCTCGTCGTGCAGTAGAAG 1500

Qy       1112 CTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAG 1171

Db       1501 AAGCGAAACGTTATGCTGAAGCAGATGATTCAGACAATGAATCTTCTTCAGAAGACTATT 1560

Qy       1172 AAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCAT 1231

Db       1561 CTGATTACAATCTAAGTTCAAGATATGCACTTGAAGCAGAAGATGAAGAAGATCGTCGTA 1620

Qy       1232 CAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATA 1291

Db       1621 ATGAAAATCGTGGTCGTGGCAAAAATAAAGTAGCGAAAGCGAAAAAAGGCGGTCGCGACG 1680

Qy       1292 ATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATG 1351

Db       1681 ATGAAAATAGCAAAAACTCGAAAAACGAGCGCGAATCTAATCGTAAAAATCAAAAAGACG 1740

Qy       1352 ATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380

Db       1741 CTAAGTTTGGTAAAGGTAAAAATGGCAAA 1769


RESULT 29
US-10-995-561-13274/c
; Sequence 13274, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13274
;   LENGTH: 415117
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(415117)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13274

  Query Match             4.0%;  Score 59.6;  DB 6;  Length 415117;
  Best Local Similarity   46.8%;  Pred. No. 1.5;
  Matches  259;  Conservative    1;  Mismatches  285;  Indels    9;  Gaps    2;

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182

Db      69341 AGASCAAGACCCTGTTGAAGAAGAAAGAAGGAAAAGAAAGGAGGAAAAGAAAGAAAGAAG 69282

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242

Db      69281 GGAAAGAAGGAAGGAAGGGAGGGGAAAAAAGAGGAAAGAAAAGAAAGAGGAAAGAAAGGA 69222

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302

Db      69221 AAGGAAAGGGAAAAGAGAAGGAAAAAG-AAAAGGGAAGGAAGGAAAGAAGGGGAAGGAAA 69163

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362

Db      69162 GGAAARAGAAAAGAGAAGGAAAATGAAAAGAGAAGGAAAAAGAAAAGGAAGGAAGGAAGG 69103

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422

Db      69102 AAAAAAGAGGAAGAAAAGGAAAGAGA--------AAAGAGAAGGAAAAAGAAAAGAGAAG 69051

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482

Db      69050 GAAGGAAGGAAAAGAGAGAGGAGAGAGGAAAGGAAGAAAAGGGAAAAGGGGAAGGTGAGG 68991

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542

Db      68990 AGAGGAGAGACGAGAAAGGAGAAAGAAGAAAGGAGGGGAAAGGGGAATGGGAAAAAGGGA 68931

Qy        543 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 602

Db      68930 AAGGGGAAAGGAAAGGGAAAAGAAGAAAAGAAGGAGGGAGGGAGGAAGGGGAGGAGGAGG 68871

Qy        603 AGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAA 662

Db      68870 AGAGGAAGAGAGAGAAGGAGGAGGAGAGGAACAGGAGGAGGAGAAGAAGAAGGAAGGAAG 68811

Qy        663 TTTAGAAAGAAAAA 676

Db      68810 AAGGAAAAGAAGAA 68797


RESULT 30
US-10-793-626-4002/c
; Sequence 4002, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4002
;   LENGTH: 3305
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4002

  Query Match             4.0%;  Score 59.4;  DB 6;  Length 3305;
  Best Local Similarity   49.9%;  Pred. No. 1.5;
  Matches  178;  Conservative    0;  Mismatches  176;  Indels    3;  Gaps    1;

Qy       1122 ATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGAT 1181

Db       3286 AAATGATAAAGATCAAAGCACACGTAAAAAAGCAGCTGAAGATATAGTTGAAAATGTAAA 3227

Qy       1182 AGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATT 1241

Db       3226 ACAAAGACAAAAAGAATTTGAAAAAGAAGAGAAGGCTCTAGATAATTCTGAAAAAGCATT 3167

Qy       1242 ATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAA 1301

Db       3166 TAAACAAGCCAAGCAATATCTTGAACATGTAGAAAACAAAGCAAAGAAAAAAGAAGTTGA 3107

Qy       1302 ACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGT 1361

Db       3106 ACAACTTGATAGTGCTATTAAAGAAAAATA---TAAATCACATGATGCTTATGCAAAGGC 3050

Qy       1362 TAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGA 1421

Db       3049 TTACAAAAAAGCACTTAATAAGGAAAAAGAACTGTTTTCTTATTTGAATGAAGATAATGC 2990

Qy       1422 AATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACT 1478

Db       2989 AACACAATCGGAAGTAGACGGAAAATCGAAAGATCTTTCTAAAGCATATAAAGAAAT 2933
```

Search completed: December 20, 2005, 23:16:05
Job time : 421.691 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:31:02 ; Search time 55.3022 Seconds
(without alignments)
5156.873 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093216_22943/app_query.fasta_1.3740
-DB=PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_179_@runat_21122005_093216_22943 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2403.5	96.6	1909	2	A45592	liver stage antige
2	382.5	15.4	1407	1	S28589	trichohyalin - rab
3	382.5	15.4	1549	1	A40691	trichohyalin - she
4	376	15.1	1898	1	A45973	trichohyalin - hum
5	359	14.4	1017	2	T15598	hypothetical prote
6	356.5	14.3	737	2	T15597	hypothetical prote
7	334.5	13.4	1192	2	A71623	probable secreted
8	326.5	13.1	1979	1	S03166	myosin heavy chain
9	323	13.0	771	1	A33430	h-caldesmon - chic
10	320.5	12.9	1039	2	S18199	myosin heavy chain
11	319.5	12.8	1166	2	H86341	hypothetical prote
12	317.5	12.8	1937	2	I38055	myosin heavy chain
13	317.5	12.8	1961	1	A61231	myosin heavy chain
14	317.5	12.8	1992	2	A47297	myosin heavy chain
15	317	12.7	1432	2	B85431	trichohyalin like
16	317	12.7	1938	2	JC5421	smooth muscle myos
17	317	12.7	1972	2	JC5420	smooth muscle myos
18	316.5	12.7	1738	2	T14867	interaptin - slime
19	316	12.7	1976	2	A59252	myosin heavy chain
20	314	12.6	1938	2	A59293	skeletal myosin he
21	313.5	12.6	2139	2	T18296	myosin heavy chain
22	313	12.6	1940	2	A29320	myosin heavy chain
23	312	12.5	1931	2	A59234	slow myosin heavy
24	312	12.5	2007	1	B43402	myosin heavy chain
25	312	12.5	2116	2	A26655	myosin heavy chain
26	310.5	12.5	678	2	A54514	glutamic acid-rich
27	310.5	12.5	1959	1	A33977	myosin heavy chain
28	310	12.5	1963	1	MWKW	myosin heavy chain
29	309.5	12.4	1956	2	T16416	hypothetical prote
30	309	12.4	1972	1	A41604	myosin heavy chain
31	307.5	12.4	665	2	B71609	hypothetical prote
32	307.5	12.4	1939	1	A46762	myosin alpha heavy
33	307.5	12.4	2017	1	A36014	myosin heavy chain
34	307.5	12.4	2057	2	S61477	myosin II heavy ch
35	306.5	12.3	1790	2	S67593	transport protein
36	306	12.3	1999	1	S21801	myosin heavy chain
37	305	12.3	746	2	T47237	myosin II heavy ch
38	305	12.3	1526	2	A45605	mature-parasite-in
39	305	12.3	1964	2	A59282	nonmuscle myosin I
40	304.5	12.2	522	2	C96608	hypothetical prote
41	304.5	12.2	1701	2	T09127	probable erythrocy
42	304	12.2	451	2	G70241	hypothetical prote
43	303.5	12.2	1938	2	I49464	alpha cardiac myos
44	303.5	12.2	1939	2	I48175	myosin heavy chain
45	302.5	12.2	1130	2	T34081	hypothetical prote

ALIGNMENTS

RESULT 1
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C;Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R;Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A;Reference number: S24597
A;Accession: S24597
A;Molecule type: DNA
A;Residues: 1-1909 <ZHU>
A;Cross-references: UNIPROT:Q25893; UNIPARC:UPI0000081753; EMBL:X56203; NID:g9915; PID:g
R;Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A;Title: Structure of Plasmodium falciparum liver stage antigen-1.
A;Reference number: A45592; MUID:92107224; PMID:1840628
A;Accession: A45592
A;Molecule type: DNA
A;Residues: 1-195;638-688;1165-1215;1590-1909 <ZH2>
A;Cross-references: UNIPARC:UPI0000177378; UNIPARC:UPI0000177379; UNIPARC:UPI000017737A;
A;Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A;Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A;Reference number: S29393; MUID:87315391; PMID:3306406
A;Accession: S29393
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 323-387 <GUE1>
A;Cross-references: UNIPARC:UPI000017737C; EMBL:M28266
R;Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A;Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by
A;Reference number: S34842
A;Accession: S34842
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 323-381,'HKAI' <GUE2>
A;Cross-references: UNIPARC:UPI000017737D; EMBL:M28266
A;Note: difference at carboxyl end due to frameshift error
C;Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: EF hand
F;154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Alignment Scores:
Pred. No.:              1.25e-119      Length:         1909
Score:                  2403.50        Matches:        486
Percent Similarity:     89.89%         Conservative:   3
Best Local Similarity:  89.34%         Mismatches:     4
Query Match:            96.64%         Indels:         51
DB:                     2              Gaps:           2

US-09-837-344-45 (1-1482) x A45592 (1-1909)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT---------------------------------- 87
              ||||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405

Qy         88 ------------------GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
                              ||||||||||||||||||||||||||||||||||||||||||||
Db       1406 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 1425

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 1445

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 1465

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 309
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db       1466 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 1485

Qy        310 GAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 369
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1486 GluGlnGlnArgAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 1505

Qy        370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 1525

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
              ||||||||||||||||   ||||||||||||||||||||   ||||||||||||||||
Db       1526 GluGlnGluArgLeuAlaAsnGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnGlu 1545

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1546 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1565

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------------ 597
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1585

Qy        597 ------------------------------------------------------------ 597

Db       1586 LeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlu 1605

Qy        598 ------------------------------GCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
                                           ||||||||||||||||||||||||||||||
Db       1606 GlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArg 1625

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 AspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGly 1645

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1646 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 1665

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1666 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 1685

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1686 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 1705

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1706 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 1725

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1726 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 1745

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 1765

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1766 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 1785

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1786 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 1805

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 1825

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1826 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 1845

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1846 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 1865

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1866 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 1885

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1886 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 1905

Qy       1468 TTTATGAAACTA 1479
              ||||||||||||
Db       1906 PheMetLysLeu 1909

RESULT 2
S28589
trichohyalin - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: S28589
R;Fietz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A;Description: Examination of the gene encoding rabbit trichohyalin.
A;Reference number: S28589
A;Accession: S28589
A;Molecule type: DNA
A;Residues: 1-1407 <FIE>
A;Cross-references: UNIPROT:P37709; UNIPARC:UPI000013738C; EMBL:Z19092; NID:g1746; PIDN:
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
```

C;Genetics:
A;Introns: 46/3
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.:	2.23e-13	Length:	1407
Score:	382.50	Matches:	127
Percent Similarity:	45.84%	Conservative:	132
Best Local Similarity:	22.48%	Mismatches:	167
Query Match:	15.38%	Indels:	139
DB:	1	Gaps:	20

US-09-837-344-45 (1-1482) x S28589 (1-1407)

```
Qy          1 CAAGAACAACAAAGCGATCTA---------GAACAAGAGAGACGTGCTAAAGAAAAGTTG 51
              |||||||||:::   :::|||         |||:::::::|||:::   :::|||:::
Db        140 GlnGluGlnGluArgGluLeuAlaGluGluGluGluGlnArgLysLysArgGluArgPhe 159

Qy         52 CAAGAA------------CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG 99
              ::::::            |||      |||   ||||||   ||||||   :::::::::
Db        160 GluGlnHisTyrSerArgGlnTyrArgAspLysGluGln---ArgLeuGlnArgGlnGlu 178

Qy        100 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 159
              |||:::|||:::::::::      |||:::|||:::|||   :::   |||   ::::::
Db        179 LeuGluGluArgArgAla------GluGluGluGlnLeuArgArgArgLysGlyArgAsp 196

Qy        160 CAACAAAGCGATCTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAA 210
                 :::      :::|||:::|||         ||||||   :::::::::|||:::
Db        197 AlaGluGluPheIleGluGluGluGlnLeuArgArgArgGluGlnGlnGluLeuLysArg 216

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG--------------TTG 255
              :::      :::   ||||||:::||||||   :::   :::                |||
Db        217 GluLeuArgGluGluGluGlnGlnArgArgGluArgArgGluGlnHisGluArgAlaLeu 236

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              ||||||::::::      |||   :::   |||      :::|||:::   :::||||||
Db        237 GlnGluGluGluGluGlnLeuLeuArgGlnArgArgTrpArgGluGluProArgGluGln 256

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
              |||           |||  ||||||      |||:::::::::|||::::::
Db        257 Gln-----------GlnLeuArgArgGluLeuGluGluIleArgGluArgGluGlnArg 272

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
              ||||||||||||   |||   :::::::::|||:::   :::|||      |||||||||
Db        273 LeuGluGlnGluGluArgArgGluGlnGlnLeuArgArgGluGlnArg---LeuGluGln 291

Qy        436 GAG--------------------AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              |||                    |||         |||:::::::::|||::::::
Db        292 GluGluArgArgGluGlnGlnLeuArgArgGluLeuGluGluIleArgGluArgGluGln 311

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
              ||||||||||||   |||   :::::::::|||:::::::::::      :::
Db        312 ArgLeuGluGlnGluGluArgArgGluGlnArgLeuGluGlnGluGluArgArgGluGln 331

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 594
              |||   :::|||      |||:::::::::|||::::::      ||||||||||||
Db        332 GlnLeuLysArgGluLeuGluGluIleArgGluArgGluGlnArgLeuGluGlnGluGlu 351

Qy        595 CGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||   ::::::   |||   |||:::   |||
Db        352 ArgArgGluGlnLeuLeuAlaGluGluValArg-------------------------- 362

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
                    |||:::   :::|||   |||:::   |||
Db        363 -----------GluGlnAlaArgGluArgGlyGluSerLeuThrArgArgTrpGlnArg 378

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
              :::||||||           |||      :::         :::   |||      |||
Db        379 GlnLeuGlu---------SerGluAlaGlyAlaArgGlnSerLysValTyrSerArgPro 395

Qy        775 ---------CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825
```

```
                      :::   ||||||   ||||||               |||      :::
Db        396 ArgArgGlnGluGluGlnSerLeuArgGlnAspGlnGluArgArgGlnArgGlnGluArg 415

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGA---------------------------- 855
              |||      :::   |||:::      |||
Db        416 GluArgGluLeuGluGluGlnAlaArgArgGlnGlnGlnTrpGlnAlaGluGluGluSer 435

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT------ 909
              |||           :::      |||      :::   :::      |||
Db        436 GluArgArgArgGlnArgLeuSerAlaArgProSerLeuArgGluArgGlnLeuArgAla 455

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAA-----------CCAGAACAAAAAGAAGATAAA 957
              ||||||:::::::::         :::               ||||||:::   ::::::
Db        456 GluGluArgGlnGluGlnGluGlnArgPheArgGluGluGluGluGlnArgArgGluArg 475

Qy        958 TCTGCTGACATACAA---------------------------------------------- 972
                  ::::::|||
Db        476 ArgGlnGluLeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGln 495

Qy        972 ---------------------------------------------------------- 972

Db        496 GlnLeuGlnGluGluAspSerPheGlnGluAspArgGluArgArgArgArgGlnGlnGlu 515

Qy        973 ----------------------------------------------------AATCAT 978
                                                                     |||
Db        516 GlnArgProGlyGlnThrTrpArgTrpGlnLeuGlnGluGluAlaGlnArgArgArgHis 535

Qy        979 ACATTA--------------GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATA 1023
              ||||||               :::      :::   :::      :::   |||
Db        536 ThrLeuTyrAlaLysProGlyGlnGlnGluGlnLeuArgGluGluGluGluLeuGlnArg 555

Qy       1024 AGTAAGTATGAGGATGAAATAAGTGCTGAATAT---GACGATTCATTAATAGATGAAGAA 1080
              |||      |||      ||||||   ::::::      :::      |||
Db        556 GluLysArgArgGlnGluArgGluArgGluTyrArgGluGluGluLysLeuGlnArgGlu 575

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||::::::                                          :::
Db        576 GluAspGluLysArg-------------------------------------ArgArg 582

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              :::|||          |||:::|||||||||:::|||   ::::::   :::
Db        583 GlnGluArgGluArgGlnTyrArgGluLeuGluGluLeuArgGlnGluGluGlnLeuArg 602

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              |||           |||:::      :::|||      ||||||:::::::::
Db        603 AspArgLysLeuArgGluGluGluGlnLeuLeuGlnGluArgGluGluGluArgLeuArg 622

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAAT------TTTAAACCAAATGATAAA 1314
              :::   ::::::      |||   ::::::               :::      ::::::
Db        623 ArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgGlnGluGluGln 642

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
              |||      |||      :::|||   :::   ::::::|||:::   :::   :::
Db        643 GluLeuArgGlnGluArgGluArgLysLeuArgGluGluGluGlnLeuLeuArgArgGlu 662

Qy       1375 GAAAAATTCATAAAA 1389
              |||:::   ::::::
Db        663 GluGlnGluLeuArg 667
```

RESULT 3
A40691
trichohyalin - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: A40691; A34209; S32633
R;Fietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A;Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-bindi
A;Reference number: A40691; MUID:93260018; PMID:7684041
A;Accession: A40691
A;Molecule type: DNA
A;Residues: 1-1549 <FIE>
A;Cross-references: UNIPROT:P22793; UNIPARC:UPI000013738D; EMBL:Z18361; NID:g295940; PIR

A;Note: sequence extracted from NCBI backbone (NCBIP:132511)
R;Fietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A;Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker
A;Reference number: A34209; MUID:90130632; PMID:2298812
A;Accession: A34209
A;Molecule type: mRNA
A;Residues: 1016-1151,1205-1257,1281-1398,'G',1400-1549 <FI2>
A;Cross-references: UNIPARC:UPI000016C5F4; GB:X51695; NID:g1827; PIDN:CAA35992.1; PID:g1
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Introns: 46/3
A;Note: single copy gene
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>
F;387-851/Region: 28-residue repeats
F;886-1519/Region: 23-residue repeats

Alignment Scores:

Pred. No.:	2.19e-13	Length:	1549
Score:	382.50	Matches:	133
Percent Similarity:	43.59%	Conservative:	129
Best Local Similarity:	22.13%	Mismatches:	194
Query Match:	15.38%	Indels:	145
DB:	1	Gaps:	18

US-09-837-344-45 (1-1482) x A40691 (1-1549)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        197 GluArgGlnArgLeuGlnValGluGlnGlnGlnArgGlnArgGlyGluLeuArgGluArg 216

Qy         61 CAAAGCGATTTA---------------GAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
Db        217 GlnGluGluValGlnLeuGlnLysArgGluThrGlnGluLeuGlnArgGluArgLeuGlu 236

Qy        106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
Db        237 GluGluGlnGlnLeuGlnLysGlnLysArgGlyLeuGluGluArgLeuLeuGluGlnGlu 256

Qy        166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db        257 ArgArgGluGlnGluLeuArgArgLysGluGlnGluArgArgGluGlnGlnLeuArgGln 276

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------------------- 264
Db        277 GluGlnGluGluAlaThrGlnGluGluIleSerGluArgGlyGluSerArgThrSerArg 296

Qy        265 ---CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAG------------------ 303
Db        297 CysGlnTrpGlnLeuGluSerGluAlaAspAlaArgGlnArgLysValTyrSerArgPro 316

Qy        304 ---TTACAAGAGCAGCAAAGC-----------------------GATTTAGAACAAGAG 336
Db        317 HisArgGlnGluGlnGlnSerArgArgGlnGluGlnGluLeuLeuGluArgGlnGlnGlu 336

Qy        337 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------------TTAGAA 381
Db        337 GlnGlnIleSerGluGluValGlnSerLeuGlnGluAspGlnGlyArgGlnArgLeuLys 356

Qy        382 CAAGAGAGACGT---------------------------------------------- 393
Db        357 GlnGluGlnArgTyrAspGlnAsnTrpArgTrpGlnLeuGluGluGluSerGlnArgArg 376

Qy        394 --------------GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
Db        377 ArgTyrThrLeuTyrAlaLysProAlaGlnArgGluGlnValArgGluGluGluGlnLeu 396

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA------ 492
Db        397 ArgLeuLysGluGluLysLeuGlnArgGluLysArgArgGlnGluArgGluArgGlnTyr 416

Qy        493 ------------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
Db        417 ArgGluValGluLeuGlnArgGluGluGluArgLeuGlnArgGluGluGluGlnLeuGln 436

Qy        535 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA--------------------------- 567
Db        437 ArgGluGluArgGluLysArgArgArgGlnGluArgGluLysGlnTyrLeuGluLysVal 456

Qy        568 ---------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAG 618
Db        457 GluLeuTrpGluGluGluGlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArg 476

Qy        619 CAGCAAAGAGATTTAGAA----------------------------CAAAGGAAGGCT 648
Db        477 GluLysGlnTyrLeuGluLysValGluLeuArgGluGluGluGlnLeuGlnArgGlnGlu 496

Qy        649 GATACGAAAAAAAATTTAGAAAGAAAAAAG---------GAACATGGAGATATATTAGCA 699
Db        497 ArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGlu 516

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
Db        517 GluGluGlnLeuGlnArgGluGlu------ArgGluLysArgArgGlnGluArgGluArg 534

Qy        760 GGATATTAT------------ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
Db        535 GlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGlu 554

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAA--------- 858
Db        555 LysArgArgGlnGluArgGluLysGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 574

Qy        859 ------TCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAA 912
Db        575 GlnLeuGlnArgGlnGluArgGlnLysArgArgGlnGluArgGluLysGlnTyrLeuGlu 594

Qy        913 GAA---------AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCT 963
Db        595 LysValGluLeuGlnGluGluGluGlnLeuGlnArgGlnGluArgGluLysArgArgGln 614

Qy        964 GACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA--- 1020
Db        615 GluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGluGlnValGlnArg 634

Qy       1021 ------------------------------------------------------ATAAGT 1026
Db        635 GlnGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGluLysGluLeuGln 654

Qy       1027 AAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
Db        655 ArgGlnGluGluArgLeuGlnGluGlu------GluGlnLeuLeuArgGluGluArgGlu 672

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
Db        673 LysArgArgGlnGluArgGluArgGlnTyrLeuGluLysValGluLeuGlnGluGluGlu 692

Qy       1147 AACATA------------------------------------GGAATTTATAAAGAA 1167
Db        693 GlnLeuGlnArgGluGluArgGluLysArgArgGlnGluArgGluArgGlnTyrLeuGlu 712

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
Db        713 LysGluGluLeuGlnArgGlnGluGluArgLeuGlnArgGluLysGluGlnLeuGlnArg 732

Qy       1228 TCATCAGAAGAA----------------------------TTATCTGAAGAAAAAATA 1257
Db        733 GluAspArgGluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGluLeuGln 752

Qy       1258 AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
Db        753 GlnGluGluAspArgLeuGlnArgGluLysGlnLeuLeuArgGluAspArgGluLysArg 772
```

```
Qy       1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
Db        773 GlnTyrLeuGluLysValGluLeuGlnArgGluGluGluGlnLeuGlnArgGluLysArg 792

Qy       1378 AAA 1380
Db        793 Arg 793
```

RESULT 4
A45973
trichohyalin - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: A45973
R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A;Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A;Reference number: A45973; MUID:93280194; PMID:7685034
A;Accession: A45973
A;Molecule type: DNA
A;Residues: 1-1898 <LEE>
A;Cross-references: UNIPROT:Q07283; UNIPARC:UPI0000013738B; GB:L09190; NID:g292835; PIDN:
A;Note: authors translated the codon AGG for residue 1714 as Pro
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Gene: GDB:THH
A;Cross-references: GDB:136223; OMIM:190370
A;Map position: 1q21-1q21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.:	4.66e-13	Length:	1898
Score:	376.00	Matches:	137
Percent Similarity:	47.39%	Conservative:	117
Best Local Similarity:	25.56%	Mismatches:	174
Query Match:	15.12%	Indels:	108
DB:	1	Gaps:	24

US-09-837-344-45 (1-1482) x A45973 (1-1898)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        522 ArgGluGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnArgLeuLysArgGln 541

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db        542 Glu---------GluGluGluArgLeuGlnGlnArgLeuArgSerGluGlnGlnLeuArg 558

Qy        121 TTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAA 177
Db        559 ArgGluGlnGluGluArgLeuGluGlnLeuLeuLysArgGluGluGluLysArgLeuGlu 578

Qy        178 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC----------------- 219
Db        579 GlnGluArgArgGluGlnArgLeuLysArgGluGlnGluGluArgArgAspGlnLeuLeu 598

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db        599 LysArgGluGluGluArgArgGlnGlnArgLeuLysArgGluGluGluGluGluArgLeuGlu 618

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
Db        619 Gln---ArgLeuLysArgGluGluValGluArg-----------LeuGluGlnGluGlu 633

Qy        340 CGTGCTAAAGAAAAGTTG----------------CAAGAACAACAAAGCGATTTA--- 378
Db        634 ArgArgAspGluArgLeuLysArgGluGluProGluGluGluArgArgHisGluLeuLeu 653

Qy        379 ---------GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429
Db        654 LysSerGluGluGlnGluGluArgArgHisGluGlnLeuArgArgGluGlnGlnGluArg 673

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA----------------CAAAGC 474
Db        674 ArgGluGlnArgLeuLysArgGluGluGluGluGluArgLeuGluGlnArgLeuLysArg 693

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------- 525
Db        694 GluHisGluGluGluArgArgGluGlnGluLeuAlaGluGluGluGlnGluGlnAlaArg 713

Qy        526 -----------------------------------GATTTAGAACAAGAG--------- 540
Db        714 GluArgIleLysSerArgIleProLysTrpGlnTrpGlnLeuGluSerGluAlaAspAla 733

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC--------------------GAT 579
Db        734 ArgGlnSerLysValLeuLeuGluAlaProGlnAlaGlyArgAlaGluAlaProGlnGlu 753

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA----------------------- 615
Db        754 GlnGluGluLysArgArgArgGluSerGluLeuGlnTrpGlnGluGluGluArgAlaHis 773

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---AAAAAAAATTTAGAAAGA 672
Db        774 ArgGlnGlnGlnGluGluGluGlnArgArgAspPheThrTrpGlnTrpGlnAlaGluGlu 793

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGT--------------- 717
Db        794 LysSerGluArgGly---------ArgGlnArgLeuSerAlaArgProProLeuArgGlu 810

Qy        718 ---------TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
Db        811 GlnArgGluArgGlnLeuArgAlaGluGluArgGlnGlnArgGluGlnArg------Phe 828

Qy        769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
Db        829 LeuProGluGluGluGluLysGluGlnArgGlyArgGlnArgArgGluArgGluLysGlu 848

Qy        829 ATGTCTATAATAGAAAAAACAAATAGA--------------------GAATCTATTACA 867
Db        849 LeuGlnPheLeuGluGluGluGluGlnLeuGlnArgArgGluArgAlaGlnGlnLeuGln 868

Qy        868 ACAAATGTTGAAGGA---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGAT 924
Db        869 GluGluGluAspGlyLeuGlnGluAspGlnGluArgArgArgGlnGluGlnArgArgAsp 888

Qy        925 GGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTA 984
Db        889 GlnLysTrpArgTrpGlnLeuGluGluGluArgLysArg------ArgArgHisThrLeu 906

Qy        985 GAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
Db        907 ----------------------------TyrAlaLysProAlaLeuGlnGluGlnLeu 916

Qy       1045 AGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAA 1104
Db        917 ArgLysGlu------GlnGlnLeuGluGlnGluGluGluGluLeuGlnArgGluGlu 934

Qy       1105 TTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
Db        935 -----------------ArgGluLysArgArgArgGlnGluGlnGluArgGlnTyrArg 948

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
Db        949 GluGluGluGlnLeuGlnGlnGluGluGluGlnLeu------LeuArgGluGluArgGlu 966

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
Db        967 LysArgArgArgGlnGluArgGluArgGlnTyrArgLysAspLysLysLeuGlnGlnLys 986

Qy       1285 AAGGATAATAATTTT---AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAA 1341
```

```
Db        987 GluGluGlnLeuLeuGlyGluGluProGluLysArgArgArgGlnGluArgGluLysLys 1006

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              |||:::   :::::::::::::   :::|||:::|||:::   ::::::
Db       1007 TyrArgGluGluGluGluLeuGlnGlnGluGluGluGlnLeuLeuArg 1022

RESULT 5
T15598
hypothetical protein C25A11.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15598
R;Favello, T.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid C25A11.
A;Reference number: Z18375
A;Accession: T15598
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1017 <FAV>
A;Cross-references: UNIPARC:UPI000017B7BA; EMBL:U39650; NID:g1049376; PID:g1049380; PIDN
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C25A11.4a
A;Introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2; 735/1; 792/3; 833/3; 873/2

Alignment Scores:
Pred. No.:              4.05e-12        Length:         1017
Score:                  359.00          Matches:        120
Percent Similarity:     47.80%          Conservative:   141
Best Local Similarity:  21.98%          Mismatches:     203
Query Match:            14.44%          Indels:         82
DB:                     2               Gaps:           18

US-09-837-344-45 (1-1482) x T15598 (1-1017)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||:::::::::::|||   :::   |||||||||   ||||||   :::      :::
Db        226 GluGluGluAlaGluLeuAlaArg---ArgArgAlaLeuGluLysGluArgIleAspArg 244

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                 |||:::::|||         |||:::|||:::   :::::::::   |||
Db        245 GluLysAlaGluGluGluArgLysThrMetGluArgLeuGluArgGluArgAlaArgLeu 264

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------CTA 174
              |||:::|||||||||   :::|||:::   |||:::::::::::::::         :::
Db        265 GluArgGluArgLeuGluGluGluArgArgGlnLysGluLysAlaGluThrGluArgIle 284

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAA 231
              |||:::|||||||||       |||:::::::::   |||:::         :::::::::
Db        285 GluArgGluArgArgGluHisGluArgIleGluIleGluArgIleGluArgIleLysArg 304

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
              ||||||      :::|||:::   :::   ::::::         |||:::|||||||||
Db        305 GluArgIleGluArgGluArgArgGluArgGluGluLysLysAlaGluGluAspArgLeu 324

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::|||:::|||:::   ::::::      :::|||:::|||||||||      |||
Db        325 LeuArgGluArgLeuGluLeuGluArg------IleGluArgGluArgArgGluLeuGlu 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACGTGCTAAAGAA 402
                  :::   |||:::   :::|||          |||   |||   |||   ::::::
Db        343 AlaArgGluArgGlnGluLeuGluLeuGlnArgArgGluAlaGluAspArgGluArgGln 362

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACTT--------- 444
              :::|||:::::::::      ::::::         |||:::||||||
Db        363 ArgLeuGluAspGluAlaArgGluMetArgArgArgGluGluGluArgArgGluAlaGlu 382

Qy        445 ----------------------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
                                         :::|||:::|||:::::::::|||
Db        383 LeuValAlaAspValHisArgGlnAlaGluGluArgGluArgLeuArgLysArgGlnGlu 402

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                   |||   ||||||      :::   :::|||::::::|||:::   |||:::|||
Db        403 ArgGluGluAlaGluArgLeuGluArgIleArgLeuGluGlnGlnLysIleAspMetGlu 422

Qy        535 CAA---------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
              :::         |||||||||   :::|||:::   :::|||::::::   :::   |||
Db        423 ArgIleAspAlaGluArgArgGluArgGluArgLysGluGluGluArgArgGluPheGlu 442

Qy        586 ------------------------CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
                                     :::|||   |||   :::::::::|||   |||
Db        443 LeuIleGluAlaAlaArgArgLysLysGluAlaArgAspArgAspArgLeuAspGluMet 462

Qy        622 CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::|||:::         |||:::   :::   ::::::      ||||||:::::::::
Db        463 GluArgGlu---ArgValArgGluGluGluGluArgArgGluLysGluArgArgGluGln 481

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
                  |||   ||||||            |||   :::            |||:::
Db        482 GluArgArgIleAlaAlaGluLysGluArgLysArgArgGlnGluGluGluGluGluIle 501

Qy        742 CCATCAGAAAATGAA--------------CGTGGATATTATATACCACATCAATCTTCT 786
                  ||||||                                        :::
Db        502 AlaArgLeuAsnGluLeuGlnArgAlaAlaAlaAlaArgGlnAlaGlnArgAsnAlaGlu 521

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAA 846
              |||   :::       :::   :::       |||   |||          :::   |||:::
Db        522 LeuAspArgGlnArgGlnArgAspGluLeuAspArgLysAlaGlnGluLeuSerGluArg 541

Qy        847 ACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACAT 906
                  ||||||                :::   |||||||||   :::   :::
Db        542 GluMetArgGlu-----------------LysGluArgArgAspArgGluArgAlaAsn 555

Qy        907 CTTGAA--------------------------GAAAAGAAAGATGGTTCAATAAAACCA 939
                 |||                         |||:::   :::      |||:::
Db        556 GluGluAlaGlnLeuAlaAspLeuLeuGluArgGluArgHisAsnGlnLeuIleArgGlu 575

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                 ::::::   :::                  |||       ||||||
Db        576 AsnGluArgArgGluAlaValGluArgAlaAsnAsnArgArgLeuGluAspArgArgSer 595

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 |||       |||      :::   :::      ::::::          |||   :::
Db        596 ArgAspLysLeuAspHisIleValArgGluArgSerGluLysGluGlnPheGluLeuGlu 615

Qy       1060 GAT---TCATTAATAGATGAAGAAGAA----------------------------GAT 1086
                   |||:::   |||:::|||
Db        616 LysArgArgLeuLeuAlaGluLysGluAlaMetAsnArgLysLysAsnHisLeuLeuSer 635

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                 |||   |||   :::             :::   |||          :::   |||
Db        636 SerGluThrLeuAlaLysLeuThrGlnProMetTyrTyrThrThrArgGluProGluVal 655

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT 1206
                   :::   :::::::::   :::   |||:::::|||      ::::::|||
Db        656 ThrThrLysValGluArgGlnValIleGluArgIleAspArgAsnValTrpValGluAsp 675

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
                 :::             :::|||:::             :::
Db        676 ValProTyrAlaProSerGlnSerAlaMetGlyTyrLeuAspAsnAsp------------ 691

Qy       1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
                 |||       :::|||       :::   |||||||||   :::
Db        692 ---------GluAsnAsnArgAspArgLeuTyrAsnProAsnAspLeuAsnArgAsnGly 708

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
                           :::|||:::      |||             |||||
Db        709 SerSerArgSerArgTyrGlnArgAlaLys--------------AsnGluLysAlaArg 723

Qy       1387 AAATCATTGTTTCATATATTT---GACGGAGACAATGAAATTTTACAGATCGTGGATGAG 1443
                 :::   :::|||      |||   :::   :::   :::      :::
Db        724 ArgAspPheTyrHisSerSerGlnAspSerAlaAspProValThrGluArgPheArgLys 743

Qy       1444 TTATCTGAAGATATAACT 1461
                 ::::::|||:::|||
```

```
Db        744 SerThrAspAspLeuThr 749

RESULT 6
T15597
hypothetical protein C25A11.4b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15597
R;Favello, T.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid C25A11.
A;Reference number: Z18375
A;Accession: T15597
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-737 <FAV>
A;Cross-references: UNIPARC:UPI000017B7BB; EMBL:U39650; NID:g1049376; PID:g1049381; PIDN
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C25A11.4b
A;Introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2

Alignment Scores:
Pred. No.:                5.78e-12        Length:         737
Score:                    356.50          Matches:        117
Percent Similarity:       47.81%          Conservative:   134
Best Local Similarity:    22.29%          Mismatches:     193
Query Match:              14.33%          Indels:         81
DB:                       2               Gaps:           17

US-09-837-344-45 (1-1482) x T15597 (1-737)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
              |||::::::::::::|||   :::   |||||||||   ||||||   :::      :::
Db        226 GluGluGluAlaGluLeuAlaArg---ArgArgAlaLeuGluLysGluArgIleAspArg 244

Qy         64 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 123
                  |||::::::|||         |||:::|||:::   :::::::::   |||
Db        245 GluLysAlaGluGluGluArgLysThrMetGluArgLeuGluArgGluArgAlaArgLeu 264

Qy        124 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------CTA 174
              |||:::|||||||||   :::|||:::   |||:::::::::::::::         :::
Db        265 GluArgGluArgLeuGluGluGluArgArgGlnLysGluLysAlaGluThrGluArgIle 284

Qy        175 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAA 231
              |||:::|||||||||      |||:::::::::  |||:::         :::::::::
Db        285 GluArgGluArgArgGluHisGluArgIleGluIleGluArgIleGluArgIleLysArg 304

Qy        232 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
              ||||||      :::|||:::   :::   ::::::         |||:::|||||||||
Db        305 GluArgIleGluArgGluArgArgGluArgGluGluLysLysAlaGluGluAspArgLeu 324

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                :::|||:::|||:::   ::::::      :::|||:::|||||||||      |||
Db        325 LeuArgGluArgLeuGluLeuGluArg------IleGluArgGluArgArgGluLeuGlu 342

Qy        352 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACGTGCTAAAGAA 402
                  :::   |||:::   :::|||          |||   |||   |||   ::::::
Db        343 AlaArgGluArgGlnGluLeuGluLeuGlnArgArgGluAlaGluAspArgGluArgGln 362

Qy        403 AAGTTGCAAGAACAACAAAGCGATTTA---------GAACAAGAGAGACTT--------- 444
              :::|||:::::::::      ::::::      |||:::||||||
Db        363 ArgLeuGluAspGluAlaArgGluMetArgArgArgGluGluGluArgArgGluAlaGlu 382

Qy        445 ----------------------------GCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
                                          :::|||:::|||:::::::::|||
Db        383 LeuValAlaAspValHisArgGlnAlaGluGluArgGluArgLeuArgLysArgGlnGlu 402

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
                  |||   ||||||      :::   :::|||::::::|||:::   |||:::|||
Db        403 ArgGluGluAlaGluArgLeuGluArgIleArgLeuGluGlnGlnLysIleAspMetGlu 422

Qy        535 CAA---------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                  :::         |||||||||   :::|||:::   :::|||::::::   :::   |||
Db        423 ArgIleAspAlaGluArgArgGluArgGluArgLysGluGluGluArgArgGluPheGlu 442

Qy        586 ------------------------CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
                                       :::|||   |||   :::::::::|||   |||
Db        443 LeuIleGluAlaAlaArgArgLysLysGluAlaArgAspArgAspArgLeuAspGluMet 462

Qy        622 CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
              :::|||:::         |||:::   :::   ::::::      ||||||:::::::::
Db        463 GluArgGlu---ArgValArgGluGluGluGluArgArgGluLysGluArgArgGluGln 481

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
                  |||   ||||||              |||   :::             |||:::
Db        482 GluArgArgIleAlaAlaGluLysGluArgLysArgArgGlnGluGluGluGluGluIle 501

Qy        742 CCATCAGAAAATGAA---------------CGTGGATATTATATACCACATCAATCTTCT 786
                  ||||||                                          :::
Db        502 AlaArgLeuAsnGluLeuGlnArgAlaAlaAlaAlaArgGlnAlaGlnArgAsnAlaGlu 521

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAA 846
              |||   :::       :::   :::       |||   |||         :::   |||:::
Db        522 LeuAspArgGlnArgGlnArgAspGluLeuAspArgLysAlaGlnGluLeuSerGluArg 541

Qy        847 ACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACAT 906
                  ||||||                  :::   |||||||||       :::   :::
Db        542 GluMetArgGlu-----------------LysGluArgArgAspArgGluArgAlaAsn 555

Qy        907 CTTGAA--------------------------GAAAAGAAAGATGGTTCAATAAAACCA 939
                  |||                         |||:::   :::      |||:::
Db        556 GluGluAlaGlnLeuAlaAspLeuLeuGluArgGluArgHisAsnGlnLeuIleArgGlu 575

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                  ::::::   :::                |||       ||||||
Db        576 AsnGluArgArgGluAlaValGluArgAlaAsnAsnArgArgLeuGluAspArgArgSer 595

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                  |||       |||       :::   :::       ::::::          |||   :::
Db        596 ArgAspLysLeuAspHisIleValArgGluArgSerGluLysGluGlnPheGluLeuGlu 615

Qy       1060 GAT---TCATTAATAGATGAAGAAGAA----------------------------GAT 1086
                  |||:::   |||:::|||
Db        616 LysArgArgLeuLeuAlaGluLysGluAlaMetAsnArgLysLysAsnHisLeuLeuSer 635

Qy       1087 GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAA 1146
                  |||   |||   :::         :::   |||         :::   |||
Db        636 SerGluThrLeuAlaLysLeuThrGlnProMetTyrTyrThrThrArgGluProGluVal 655

Qy       1147 AACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT 1206
                  :::   :::::::::   :::   |||:::::::|||       :::::::|||
Db        656 ThrThrLysValGluArgGlnValIleGluArgIleAspArgAsnValTrpValGluAsp 675

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
                  :::         :::|||:::         :::
Db        676 ValProTyrAlaProSerGlnSerAlaMetGlyTyrLeuAspAsnAsp------------ 691

Qy       1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
                  |||       :::|||       :::   |||||||||   :::
Db        692 ---------GluAsnAsnArgAspArgLeuTyrAsnProAsnAspLeuAsnArgAsnGly 708

Qy       1327 GAGCATATTAAAAAATATAAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATA 1386
                  :::|||:::       |||                     ||||||
Db        709 SerSerArgSerArgTyrGlnArgAlaLys--------------AsnGluLysAlaArg 723

Qy       1387 AAATCATTGTTTCAT 1401
                  :::   :::|||
Db        724 ArgAspPheTyrHis 728

RESULT 7
A71623
probable secreted protein PFB0115w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: A71623
```

```
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71623
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1192 <GAR>
A;Cross-references: UNIPROT:O96127; UNIPARC:UPI0000076661; GB:AE001373; GB:AE001362; NID
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0115w

Alignment Scores:
Pred. No.:                7.64e-11      Length:          1192
Score:                    334.50        Matches:          112
Percent Similarity:       46.85%        Conservative:     126
Best Local Similarity:    22.05%        Mismatches:       193
Query Match:              13.45%        Indels:            77
DB:                       2             Gaps:              16

US-09-837-344-45 (1-1482) x A71623 (1-1192)

Qy         10 CAAAGCGATCTAGAACAAGAGAGA---------------------------------CGT 36
              ||||||      :::|||||||||                                   :::
Db        296 GlnSerLysTyrLysGlnGluArgIleGluIleLeuAspAspAsnGlyLysGluLeuLys 315

Qy         37 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA---------GATAGACTT 87
              :::   :::   ::::::|||::::::      :::|||:::               :::
Db        316 SerHisLysAsnIleLysGluGluLysGlyGlyIleGluLysThrAspThrThrAsnIle 335

Qy         88 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA 147
              |||      |||:::::::::::::::   :::   :::   |||:::      ::::::
Db        336 AlaAspIleLysIleLysLysGluGluArgGluThrLysAspGluLys------GluLys 353

Qy        148 AAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 207
                 :::|||:::         |||::::::      :::      :::|||
Db        354 AsnIleGlnGlnLeuValLysAspValGlnLeuIleLysValGlyGluGluThrLysAsp 373

Qy        208 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 267
              :::::::::   |||   |||               ::::::         |||:::
Db        374 AspGluLysGluAspLysGluGlyThrAspAspGluGluAspThrAspAspGluGluAsp 393

Qy        268 AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA 327
              :::|||   |||:::|||         :::|||         ::::::::::::
Db        394 ThrAspAspGluGluAspThrAspAspGluGluAspThrSerAspGluGluThrThrGly 413

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
              :::||||||   :::   :::   ::::::   |||::::::::::::
Db        414 AspGlnGluAsnLysGluGluThrGluValAspGluLysLysThrGlu------------ 429

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
                 :::|||:::|||:::|||:::|||   :::   :::   |||:::::::::
Db        430 ---LysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGluGlu 448

Qy        448 AAAGAAAAG---TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 504
                 ||||||      :::|||   :::   |||   |||:::      :::   ||||||
Db        449 SerGluLysAspLysGluGluSerGluLysAspLysGluGluSerGluLysAspLysGlu 468

Qy        505 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 564
              |||   :::|||   :::          |||   |||:::      :::          :::
Db        469 LysThrGluGluAspGluGluLysThrGluAspGluLysGlyThrGluValTyrLysLys 488

Qy        565 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAA 624
              |||          |||   ::::::|||:::          :::      :::   :::
Db        489 GluThrAspValAspGluLysLysGluLysGlyGluTyrGlyGluGlyThrAspAspGlu 508

Qy        625 AGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACAT 684
                 |||   |||:::   :::   |||   :::      :::|||   ||||||
Db        509 GluAspLysGluLysGluGluAspAspGluGluThrLysValGluGluLysLysThrGlu 528

Qy        685 GGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCA 744
                |||
Db        529 LysAsp---------------------------------------------------- 530

Qy        745 TCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA 804
                  |||   |||                   :::         :::|||
Db        531 ---------GluGluGly-----------------ThrAspTyrGluGluAspThrAsp 541

Qy        805 GGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA---GAATCT 861
                 :::   :::|||   :::                 :::||||||   |||   |||
Db        542 AspSerAspLysAspGluGluThrLysValGluGluLysLysThrGluArgAspGluGlu 561

Qy        862 ATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA 921
                 |||         |||         ::::::   |||         |||:::   :::
Db        562 GluThrGluGluAspGluLysGluThrGluValGluLysLysLysThrGluLysAspGlu 581

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                 :::|||:::         |||      :::|||         |||::::::
Db        582 GluGlyThrAspTyrGluGluAspThrAspAspSerAspLysAspValGluThrGluVal 601

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG------TATGAG 1035
                 ||||||         |||      |||   :::   ::::::         :::
Db        602 GluGluThr------------AspAlaGluAspLysGluGluAsnGluGluGlyThrAsp 617

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT---GATGAA 1092
                 ||||||              :::      |||   |||:::|||||||||   :::|||
Db        618 AspGluGluAspLysValGluGluThrAspLeuAspAspGlnGluGluAspGlyGluGlu 637

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
                 |||   ::::::   |||          :::   |||:::   :::|||::::::
Db        638 AspLysGluAspAspLysGluLysAspLysGluAspAspLysGluAspAspLysGlu--- 656

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA--------------------AAT 1191
                 |||:::   ||||||      ||||||                         :::
Db        657 ---------LysAspLysGluAspAspLysGluLysTyrLysGluAspAspLysGluAsp 673

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
                 :::      ||||||   ::::::      |||   :::   |||:::      |||:::
Db        674 AspLysGluAspAspLysGluLysAspLysGluAspAsnLysGluLysAspLysGluAsp 693

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGAT 1311
                 :::|||   |||:::   :::|||   ||||||         |||   :::|||
Db        694 AsnLysGluLysAspLysGluAspAspLysGluLysAsp---------LysGluAspAsp 710

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAA 1371
                 |||         |||      |||      :::   ||||||:::   |||||||||
Db        711 LysGluLysAspLysGluAspAsnLys------GluLysAspLysGluAspAsnLysGlu 728

Qy       1372 AAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
                 |||:::|||                  |||            |||      :::   :::
Db        729 LysAspLysGluAspAspLysGluLysHis-----------AspLysHisValArgArg 744

Qy       1432 ATCGTGGATGAGTTATCTGAAGAT 1455
                 |||      ::::::   :::|||
Db        745 IleLysLysLysMetLysAspAsp 752

RESULT 8
S03166
myosin heavy chain, gizzard smooth muscle [similarity] - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S03166; A27066; A26045; A36604; A43298
R;Yanagisawa, M.; Hamada, Y.; Katsuragawa, Y.; Imamura, M.; Mikawa, T.; Masaki, T.
J. Mol. Biol. 198, 143-157, 1987
A;Title: Complete primary structure of vertebrate smooth muscle myosin heavy chain deduc
A;Reference number: S03166; MUID:88118918; PMID:2892941
A;Accession: S03166
A;Molecule type: mRNA
A;Residues: 1-1979 <YAN>
A;Cross-references: UNIPROT:P10587; UNIPARC:UPI0000171370; EMBL:X06546; NID:g63633; PIDN
A;Note: part of this sequence was confirmed by protein sequencing
R;Maita, T.; Onishi, H.; Yajima, E.; Matsuda, G.
J. Biochem. 102, 133-145, 1987
```

```
A;Title: Amino acid sequence of the amino-terminal 24 kDa fragment of the heavy chain of
A;Reference number: A27066; MUID:88032919; PMID:3312184
A;Accession: A27066
A;Molecule type: protein
A;Residues: 2,'Z',4-204 <MAI>
A;Cross-references: UNIPARC:UPI0000173E2B
R;Onishi, H.; Maita, T.; Miyanishi, T.; Watanabe, S.; Matsuda, G.
J. Biochem. 100, 1433-1447, 1986
A;Title: Amino acid sequence of the 203-residue fragment of the heavy chain of chicken g
A;Reference number: A26045; MUID:87194651; PMID:3571180
A;Accession: A26045
A;Molecule type: protein
A;Residues: 653-855 <ONI>
A;Cross-references: UNIPARC:UPI0000173E2C
R;Onishi, H.; Maita, T.; Matsuda, G.; Fujiwara, K.
J. Biol. Chem. 265, 19362-19368, 1990
A;Title: Lys-65 and Glu-168 are the residues for carbodiimide-catalyzed cross-linking be
A;Reference number: A36604; MUID:91035476; PMID:1977747
A;Accession: A36604
A;Status: preliminary
A;Molecule type: protein
A;Residues: 54-67;146-183 <ON2>
A;Cross-references: UNIPARC:UPI0000173E2D; UNIPARC:UPI0000173E2E
R;Cole, D.G.; Yount, R.G.
Biochemistry 31, 6186-6192, 1992
A;Title: Stability and photochemical properties of vanadate-trapped nucleotide complexes
A;Reference number: A43298; MUID:92329440; PMID:1385724
A;Accession: A43298
A;Status: preliminary
A;Molecule type: protein
A;Residues: 169-183 <COL>
A;Cross-references: UNIPARC:UPI0000173E2F
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; blocked amino end; coiled coil; hydrolase; methylated am
F;87-777/Domain: myosin motor domain homology <MMOT>
F;177-184/Region: nucleotide-binding motif A (P-loop)
F;565-578/Region: actin binding #status predicted
F;639-653/Region: actin binding #status predicted
F;850-1940/Domain: coiled coil <COI>
F;850-1290/Region: S2
F;1291-1979/Region: light meromyosin
F;1941-1979/Domain: carboxyl-terminal <CBT>
F;2/Modified site: blocked amino end (Ser) (in mature form) #status experimental
F;128/Modified site: N6,N6,N6-trimethyllysine (Lys) #status experimental
F;183/Binding site: ATP (Lys) #status predicted
F;707,717/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                1.85e-10        Length:          1979
Score:                    326.50          Matches:         154
Percent Similarity:       42.17%          Conservative:    134
Best Local Similarity:    22.55%          Mismatches:      186
Query Match:              13.13%          Indels:          209
DB:                       1               Gaps:            35

US-09-837-344-45 (1-1482) x S03166 (1-1979)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::|||::::::         :::|||   :::      |||:::|||:::|||:::|||
Db        938 GluGluGluGluGluArgSerGlnGlnLeuGlnAlaGluLysLysLysMetGlnGlnGln 957

Qy         61 CAAAGCGAT------------TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA--- 105
                  |||            ||||||::::::      |||::::::|||||||||
Db        958 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 977

Qy        106 ----------------------------GAGCAGCAAAGCGATTTA------------ 123
                                          :::   :::   |||:::
Db        978 GluLysValThrAlaAspGlyLysIleLysLysMetGluAspAspIleLeuIleMetGlu 997

Qy        124 ---GAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
                  :::      :::|||  ||||||:::      |||:::|||:::  |||||||||
Db        998 AspGlnAsnAsnLysLeuThrLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1017

Qy        175 --------------GAACAAGAGAGACGTGCTAAAGAA---------------------- 198
                             |||:::|||   :::||||||
Db       1018 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1037

Qy        199 ----------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAA 228
                                         :::|||::::::::::::         :::
Db       1038 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1057

Qy        229 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT--- 285
                :::      :::  |||  ||||||:::  :::  |||||||||  :::
Db       1058 GluLeuGluLysIleLysArgLysLeuGluGlyGluSerSerAspLeuHisGluGlnIle 1077

Qy        286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
                                                :::|||||||||      |||:::
Db       1078 AlaGluLeuGlnAlaGlnIleAlaGluLeuLysAlaGlnLeuAlaLysLysGluGluGlu 1097

Qy        304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG---------------------------- 336
                ||||||        :::  ||||||  |||
Db       1098 LeuGlnAlaAlaLeuAlaArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLys 1117

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
                    |||      :::      :::  :::  |||  |||||||||  |||:::
Db       1118 LysIleArgGluLeuGluSerHisIleSerAspLeuGlnGluAspLeuGluSerGluLys 1137

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||  ::::::|||:::  ||||||  :::|||     |||
Db       1138 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAlaLeu 1157

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
                |||   :::|||::::::          :::      :::||||||  |||||||||
Db       1158 LysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLys 1177

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                          |||:::|||:::            |||
Db       1178 ArgGluGlnGluValThrValLeuLysArgAlaLeuGluGluGluThrArgThrHisGlu 1197

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||  |||  |||      :::  ::::::|||
Db       1198 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1217

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                |||||||||  :::||||||||||      |||  :::  ::::::  ||||||:::
Db       1218 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLys 1237

Qy        640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
                ||||||                                |||::::::::|||
Db       1238 AspAsnAlaAspLeuAlaAsnGluIleArgSerLeuSerGlnAlaLysGlnAspValGlu 1257

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA--------- 720
                ||||||:::      ::::::      :::||||||      :::
Db       1258 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysTyrSerAspGly 1277

Qy        721 --------------------------GAAATACCAGCTATAGAACTTCCATCAGAAAAT 753
                                        ::::::      :::::::::      ||||||
Db       1278 GluArgValArgThrGluLeuAsnGluLysValHisLysLeuGlnIleGluValGluAsn 1297

Qy        754 GAACGTGGATATTATATACCACATCAATCT---------TCTTTACCTCAGGAC------ 798
                                  :::|||            |||   :::|||
Db       1298 ValThrSerLeuLeuAsnGluAlaGluSerLysAsnIleLysLeuThrLysAspValAla 1317

Qy        799 AACAGAGGGAATAGT---AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA 855
                |||:::     :::|||::::::|||:::       |||        :::
Db       1318 ThrLeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgGlnLys 1337

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
                ::::::||||||  :::          |||      |||      |||:::|||
Db       1338 LeuAsnValThrThrLysLeuArgGlnLeuGluAsp---AspLysAsnSerLeuGlnGlu 1356

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
                :::  |||      ::::::  :::|||                    :::::::::
Db       1357 GlnLeuAspGluGluValGluAlaLysGln-----------------AsnLeuGluArg 1370

Qy        976 CAT------ACATTAGAGACAGTAAATATTTCTGAT-----------GTTAATGATTTT 1017
```

```
Db       1371 HisIleSerThrLeu---ThrIleGlnLeuSerAspSerLysLysLysLeuGlnGluPhe 1389

Qy       1018 ------------------------------------------------------------CAA 1020

Db       1390 ThrAlaThrValGluThrMetGluGluGlyLysLysLysLeuGlnArgGluIleGluSer 1409

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC--------------GATTCA 1065

Db       1410 LeuThrGlnGlnPheGluGluLysAlaAlaSerTyrAspLysLeuGluLysThrLysAsn 1429

Qy       1066 TTAATAGATGAAGAAGAAGATGAT------GAAGACTTAGACGAATTTAAGCCTATTGTG 1119

Db       1430 ArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGlnArgGlnLeuVal 1449

Qy       1120 ----------------------CAATATGACAATTTCCAAGATGAAGAAAAC------ 1149

Db       1450 SerAsnLeuGluLysLysGlnLysLysPheAspGlnMetLeuAlaGluGluLysAsnIle 1469

Qy       1149 ------------------------------------------------------------ 1149

Db       1470 SerSerLysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThr 1489

Qy       1150 ------ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203

Db       1490 LysAlaLeuSerLeuAlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeuGlu 1509

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263

Db       1510 ArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLysAspAspVal 1529

Qy       1264 GGAAAG---------AAATATGAAAAAACAAAGGAT-----------------AATAAT 1296

Db       1530 GlyLysAsnValHisGluLeuGluLysSerLysArgThrLeuGluGlnGlnValGluGlu 1549

Qy       1297 TTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAG 1356

Db       1550 MetLysThrGlnLeuGluGluLeuGluAspGlu---LeuGlnAlaAlaGluAspAlaLys 1568

Qy       1357 ---------CAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407

Db       1569 LeuArgLeuGluValAsnMetGlnAla---------MetLysSerGlnPhe--------- 1582

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467

Db       1583 ---------GluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArgArgGln 1599

Qy       1468 TTTATGAAA 1476

Db       1600 LeuLeuLys 1602
```

RESULT 9
A33430
h-caldesmon - chicken
N;Alternate names: caldesmon, smooth muscle; calmodulin- and actin-binding protein
C;Species: Gallus gallus (chicken)
C;Date: 27-Feb-1990 #sequence_revision 27-Jun-1994 #text_change 09-Jul-2004
C;Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022
R;Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 164, 503-511, 1989
A;Title: Primary structure and functional expression of h-caldesmon complementary DNA.
A;Reference number: A33430; MUID:90026426; PMID:2803315
A;Accession: A33430
A;Molecule type: mRNA
A;Residues: 1-771 <HAY>
A;Cross-references: UNIPROT:P12957; UNIPARC:UPI0000126E45; GB:M28417; NID:g211895; PIDN:
A;Experimental source: gizzard
A;Note: part of this sequence was confirmed by protein sequencing
R;Bryan, J.; Imai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.
J. Biol. Chem. 264, 13873-13879, 1989
A;Title: Cloning and expression of a smooth muscle caldesmon.
A;Reference number: A32642; MUID:89340480; PMID:2760048
A;Accession: A32642
A;Molecule type: mRNA
A;Residues: 1-318,334-771 <BRY>
A;Cross-references: UNIPARC:UPI00001713AB; GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:
A;Note: the authors translated the codon GAA for residue 743 as Leu
A;Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)
R;Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 161, 38-45, 1989
A;Title: 35kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo
A;Reference number: A32445; MUID:89273666; PMID:2730665
A;Accession: A32445
A;Molecule type: mRNA
A;Residues: 466-771 <HA2>
A;Cross-references: UNIPARC:UPI000015624D; GB:M26684; NID:g211897; PIDN:AAA48811.1; PID:
R;Mak, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.
J. Biol. Chem. 266, 19971-19975, 1991
A;Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphoryla
A;Reference number: A41064; MUID:92041815; PMID:1939059
A;Accession: A41064
A;Molecule type: protein
A;Residues: 597-600;678-696;711-721 <MAK>
A;Cross-references: UNIPARC:UPI0000173E63; UNIPARC:UPI0000173E64; UNIPARC:UPI0000173E65
R;Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.
Biochem. Biophys. Res. Commun. 162, 746-752, 1989
A;Title: Caldesmon has two calmodulin-binding domains.
A;Reference number: A60461; MUID:89334885; PMID:2757638
A;Accession: A60461
A;Molecule type: protein
A;Residues: 2-17,'X',19-38;466-485 <WAN>
A;Cross-references: UNIPARC:UPI0000173E66; UNIPARC:UPI0000173E67
R;Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.
Biochem. Biophys. Res. Commun. 197, 145-153, 1993
A;Title: Common structural and expressional properties of vertebrate caldesmon genes.
A;Reference number: PC2003; MUID:94071934; PMID:8250919
A;Accession: PC2003
A;Molecule type: DNA
A;Residues: 74-419 <HAR>
A;Cross-references: UNIPARC:UPI0000173E68
R;Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.
J. Biochem. 106, 778-783, 1989
A;Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon w
A;Reference number: PX0022; MUID:90130380; PMID:2613684
A;Accession: PX0022
A;Molecule type: protein
A;Residues: 462-477,'D',479-563;674-762,'A',763-771 <TAK>
A;Cross-references: UNIPARC:UPI0000173E69; UNIPARC:UPI0000173E6A
C;Comment: This protein plays a vital role in the regulation of smooth muscle and nonmus
C;Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C;Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C;Superfamily: caldesmon
C;Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; muscl
F;1-771/Product: h-caldesmon #status predicted <HMAT>
F;1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>
F;266-390/Region: 13-residue repeats
F;511-582/Region: tropomyosin binding
F;622-636/Region: tropomyosin binding
F;597,682,717/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status experime
F;688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimental

Alignment Scores:

Pred. No.:	3.31e-10	Length:	771
Score:	323.00	Matches:	122
Percent Similarity:	47.07%	Conservative:	111
Best Local Similarity:	24.65%	Mismatches:	172
Query Match:	12.99%	Indels:	90
DB:	1	Gaps:	21

US-09-837-344-45 (1-1482) x A33430 (1-771)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db        179 LysLysGluGluLysAspSerGluGluGluLysProLysGluValProThrGluGluAsn 198

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
```

```
Db        199 GlnValAsp--------------ValAlaValGluLysSerThrAspLysGluGluVal 213

Qy        121 TTAGAACAAGAGAGACTTGCT--------------------AAAGAAAAGTTGCAAGAA 159
Db        214 ValGluThrLysThrLeuAlaValAsnAlaGluAsnAspThrAsnAlaMetLeuGluGly 233

Qy        160 CAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 219
Db        234 GluGlnSerIleThrAspAlaAlaAspLysGluLysGluGluAlaGluLysGluArgGlu 253

Qy        220 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 279
Db        254 LysLeuGluAlaGluGluLysGluArgLeuLysAlaGluGluGluLysLysAlaAlaGlu 273

Qy        280 CAAGATAGACTTGCTAAAGAAAAGTTA------------CAAGAGCAGCAAAGCGATTTA 327
Db        274 GluLysGlnLysAlaGluGluGluLysLysAlaAlaGluGluArgGluArgAlaLysAla 293

Qy        328 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 387
Db        294 GluGluGluLysArgAlaAlaGlu------GluArgGluArgAlaLysAlaGluGluGlu 311

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
Db        312 ArgLysAlaAlaGlu------GluArgGluArgAlaLysAlaGluGluGluArgLysAla 329

Qy        448 AAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 498
Db        330 AlaGluGluArgAlaLysAlaGluGluGluArgLysAlaAlaGluGluArgAlaLysAla 349

Qy        499 AAAGAA------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
Db        350 GluGluGluArgLysAlaAlaGluGluArgAlaLysAlaGluLysGluArgLysAlaAla 369

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT------------ 600
Db        370 Glu------GluArgGluArgAlaLysAlaGluGluGluLysArgAlaAlaGluGluLys 387

Qy        601 ------------AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db        388 AlaArgLeuGluAlaGluLysLeuLysGluLysLysLys---MetGluGluLysLysAla 406

Qy        649 GATACGAAAAAA---------AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
Db        407 GlnGluGluLysAlaGlnAlaAsnLeuLeuArgLysGlnGlu------------------ 420

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
Db        421 GluAspLysGluAlaLysValGluAlaLysLysGluSerLeuProGluLysLeuGln--- 439

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
Db        440 ----------------------------ProThrSerLysLysAspGlnValLysAsp 449

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
Db        450 AsnLysAspLys---------GluLysAlaProLysGluGluMetLysSerValTrpAsp 466

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
Db        467 ArgLysArgGlyVal----------------------------------------Pro 472

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
Db        473 GluGlnLys------------AlaGlnAsnGlyGluArgGluLeuThrThrProLysLeu 488

Qy       1000 ---TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
Db        489 LysSerThrGluAsnAlaPheGlyArgSerAsnLeuLysGlyAlaAlaAsnAlaGluAla 508

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAA------GACTTAGACGAATTTAAG 1110
```

```
Db        509 GlySerGluLysLeuLysGluLysGlnGlnGluAlaAlaValGluLeuAspGluLeuLys 528

Qy       1111 CCTATTGTGCAATATGACAAT------TTCCAAGATGAAGAAAACATAGGAATTTATAAA 1164
Db        529 LysArgArgGluGluArgArgLysIleLeuGluGluGluGluGlnLysLysLysGlnGlu 548

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
Db        549 GluAlaGluArgLysIleArgGluGluGluGluLysLysArgMetLysGluGluIleGlu 568

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA---------AAAGGAAAGAAATAT 1275
Db        569 ArgArgArgAlaGluAlaAlaGluLysArgGlnLysValProGluAspGlyValSerGlu 588

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATT 1335
Db        589 GluLysLysProPheLysCysPheSerProLysGlySerSerLeuLysIleGluGluArg 608

Qy       1336 AAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db        609 AlaGluPhe--------------LeuAsnLysSerAlaGlnLys 618
```

RESULT 10
S18199
myosin heavy chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 02-Feb-2001
C;Accession: S18199
R;Stewart, A.F.R.; Camoretti-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.
J. Mol. Evol. 33, 357-366, 1991
A;Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy c
A;Reference number: S18199; MUID:92130260; PMID:1774788
A;Accession: S18199
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1039 <STE>
A;Cross-references: UNIPARC:UPI0000171256; EMBL:X59552; NID:g62995; PIDN:CAA42130.1; PI
A;Note: in the authors' translation 45-Lys is shown after residue 40, and, consequently
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: ATP; nucleotide binding; P-loop
F;402-409/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:

Pred. No.:	4.26e-10	Length:	1039
Score:	320.50	Matches:	142
Percent Similarity:	40.99%	Conservative:	115
Best Local Similarity:	22.65%	Mismatches:	187
Query Match:	12.89%	Indels:	183
DB:	2	Gaps:	27

US-09-837-344-45 (1-1482) x S18199 (1-1039)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA--------------- 45
Db         32 GluGluMetAsnSerGluLeuThrSerLysLysArgLysLeuGluAspGluPheAlaLeu 51

Qy         46 AAGTTGCAAGAACAACAAAGCGATTTAGAA------CAAGATAGACTTGCTAAAGAAAAG 99
Db         52 GluLeuLysLysAspIleAspAspLeuGluIleThrLeuAlaLysValGluLysGluLys 71

Qy        100 --------------------------TTACAAGAGCAGCAAAGCGATTTAGAACAA--- 129
Db         72 GlnAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaThrLeuAspGluAsn 91

Qy        130 ---GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAA------AGC 168
Db         92 IleSerLysLeuThrLysGluLysLysSerLeuGlnGluAlaHisGlnGlnValLeuAsp 111

Qy        169 GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAGTTGCAA 207
Db        112 AspLeuGlnAlaGluGluAspLysValAsnArgLeuSerLysAlaLysValLysLeuGlu 131

Qy        208 GAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------------ 237
```

```
Db        132 GlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysValArgMetAspLeu 151

Qy        238 ---CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAAAGCGAT 273

Db        152 GluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuThrGlnGluSerValMetAsp 171

Qy        274 TTAGAACAAGATAGACTT---GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA 330

Db        172 LeuGluAsnAspLysLeuGlnMetGluGluLysLeuLysLysLysGluPheGluMetSer 191

Qy        331 CAA-----------------------------------------------GAGAGACGTGCT 345

Db        192 GlnLeuAsnSerLysIleGluAspGluGlnAlaIleValMetGlnLeuGlnLysLysIle 211

Qy        346 AAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 396

Db        212 LysGluLeuGlnAlaArgIleGluGluLeuGluGluGluLeuGluAlaGluArgAlaAla 231

Qy        397 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAG 456

Db        232 ArgAlaLysValGluLysGlnArgSerAspLeuAlaArgGlu---------LeuGluGlu 248

Qy        457 TTGCAAGAACAA----------------------CAAAGCGATTTAGAACAAGAGAGA 492

Db        249 LeuSerGluArgLeuGluGlyAlaGlyGlyAlaThrAlaAlaGlnLeuGluMetAsnLys 268

Qy        493 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA--------------- 537

Db        269 LysArgGluAlaGluPheLeuLysLeuAlaArgAspLeuGluGluAlaThrLeuHisTyr 288

Qy        538 --------------------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 576

Db        289 GluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspArgArgGlyGluMetGlyGlu 308

Qy        577 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAA 636

Db        309 GlnLeuAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLys 328

Qy        637 CAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTA 696

Db        329 MetGluValAspAspLeuThrAlaAsnMetGluGlnThrValLys---GlyLysAlaAsn 347

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAA------------------------------ 723

Db        348 AlaGluLysLeuTrpGlyThrTyrGluAspHisLeuAsnGluThrLysThrLysLeuAsp 367

Qy        724 ------------------------ATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759

Db        368 GluMetThrArgLeuMetAsnAspLeuThrThrGlnLysThrLysLeuGlnSerGluAsn 387

Qy        760 GGATATTATATA------------CCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807

Db        388 GlyGluPheValArgGlnLeuGluGluLysGluSerLeuIleSerGlnLeuSerArgGly 407

Qy        808 AATAGTAGAGATTCCAAGGAAATG---------------------------------- 831

Db        408 LysThrSerPheThrGlnGlnIleGluGluLeuArgArgGlnLeuGluGluGluThrLys 427

Qy        832 ------------------------------------------------TCTATAATAGAA 843

Db        428 SerLysAsnAlaLeuAlaHisAlaLeuGlnAlaAlaArgHisAspCysAspLeuLeuArg 447

Qy        844 AAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGA 903

Db        448 GluGlnTyrGluGluGluGlnGluAlaLysAlaGluLeuGlnArgAlaLeuSerLysGly 467

Qy        904 CATCTTGAA--------------------GAAAAGAAAGATGGTTCAATAAAACCAGAA 942

Db        468 AsnAlaGluValAlaGlnTrpArgThrLysTyrGluThrAspAlaIleGlnArgThrGlu 487

Qy        943 CAAAAAGAAGATAAATCTGCTGACATA---------CAAAATCATACATTAGAGACAGTA 993

Db        488 GluLeuGluAspAlaLysLysLysLeuLeuAlaArgLeuGlnGluAlaGluGluAlaIle 507

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053

Db        508 GluAlaAlaAsnAlaLysCysSerSerLeuGluLysAlaLysHisArgLeuGlnAsnGlu 527

Qy       1054 TATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCT 1113

Db        528 GlnGluAspMetMetIleAspLeuGluLysAlaAsnSerAlaAlaAlaSerLeu------ 545

Qy       1114 ATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA 1173

Db        546 ----------------------AspLysLysGlnArgGlyPheAspLysIleIleAsn 557

Qy       1174 GATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCA 1233

Db        558 AspTrpLysGlnLysTyrGluGluSerGlnAlaGluLeuGluAlaSerGlnLysGluAla 577

Qy       1234 GAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAAT 1293

Db        578 ArgSerLeuSerThrGluLeuPheLysLeuLysAsnAlaTyrGluGluThrLeuAspHis 597

Qy       1294 ------AATTTTAAACCAAATGATAAAAGTTTGTATGATGAG----------------- 1329

Db        598 LeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGluGluIleSerAspLeuThrAsn 617

Qy       1330 ----------------------------CATATTAAAAAATATAAAAATGATAAGCAG 1359

Db        618 GlnIleSerGluGlyAsnLysAsnLeuHisGluIleGluLysValLys------LysGln 635

Qy       1360 GTTAATAAGGAAAAGGAAAAA 1380

Db        636 ValGluGlnGluLysSerGlu 642
```

RESULT 11
H86341
hypothetical protein F9H16.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: H86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86341
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1166 <STO>
A;Cross-references: UNIPROT:Q9SYP6; UNIPARC:UPI00000A7C33; GB:AE005172; NID:g4836896; PI
C;Genetics:
A;Map position: 1

Alignment Scores:

Pred. No.:	4.71e-10	Length:	1166
Score:	319.50	Matches:	136
Percent Similarity:	44.24%	Conservative:	110
Best Local Similarity:	24.46%	Mismatches:	189
Query Match:	12.85%	Indels:	121
DB:	2	Gaps:	26

US-09-837-344-45 (1-1482) x H86341 (1-1166)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63

Db         17 GluGluGluAlaAspLeuLysLysSerArgArgAspArgAspArgSerAsnGluArgLys 36
```

```
Qy        64 AGCGAT-----------------TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
Db        37 LysAspLysGlySerGluLysArgArgGluLysAspArg---ArgLysLysArgValLys 55

Qy       106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
Db        56 SerSerAspSerGluAspAspTyrAspArgAspAspAspGluGlu---ArgGluLysArg 74

Qy       166 AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 225
Db        75 LysGluLysGluArgGluArgArgArgArgAspLys-----------------AspArg 88

Qy       226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA---GAACAA 282
Db        89 ValLysArgArgSerGluArgArgLysSerSerAspSerGluAspAspValGluGluGlu 108

Qy       283 GATAGACTTGCTAAAGAAAAGTTACAAGAG---CAGCAAAGCGATTTAGAACAAGAGAGA 339
Db       109 AspGluArgAspLysArgArgValAsnGluLysGluArgGlyHisArgGluHisGluArg 128

Qy       340 --------------CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 384
Db       129 AspArgGlyLysAspArgLysArgAspArgGluArgGluGluArgLysAspLysGluArg 148

Qy       385 GAGAGA---CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
Db       149 GluArgGluLysAspArgGluArgArgGluArgGluArgGluGluArgGluLysGluArg 168

Qy       442 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
Db       169 Val---LysGluArgGluArgArgGluArgGluAspGlyGluArgAspArgArgGluArg 187

Qy       502 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 561
Db       188 GluLys------GluArgGlySerArgArgAsnArgGluArgGluArgSerArgGluVal 205

Qy       562 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
Db       206 GlyAsnGluGluSerAspAspAspValLysArgAspLeuLysArgArgArgLysGluGly 225

Qy       622 CAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAA 681
Db       226 Gly---------GluArgLysGluLysGluArgGluLysSerValGlyArgSerSerArg 242

Qy       682 CATGGAGATATA------------TTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db       243 HisGluAspSerProLysArgLysSerValGluAspAsnGlyGluLysLysGluLysLys 262

Qy       730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db       263 ThrArgGluGluGluLeuGluAspGlu--------------------GlnLysLysLeu 275

Qy       790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
Db       276 AspGluGluValGluLysArgArgArgArgValGlnGluTrpGlnGluLeuLysArgLys 295

Qy       850 AATAGAGAATCTATTACAACAAAT---------------------------------- 873
Db       296 LysGluGluAlaGluSerGluSerLysGlyAspAlaAspGlyAsnGluProLysAlaGly 315

Qy       874 ------------GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA 921
Db       316 LysAlaTrpThrLeuGluGlyGluSerAspAspGluGluGlyHisProGluGluLysSer 335

Qy       922 ------------------GATGGTTCAATAAAACCAGAACAAAAAGAAGAT---AAATCT 960
Db       336 GluThrGluMetAspValAspGluGluThrLysProGluAsnAspGlyAspAlaLysMet 355

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db       356 ValAspLeuGluAsnGluThrAlaAlaThrValSerGluSerGly--------------- 370

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
```

```
Db       371 ----------------------------------GlyAspGlyAlaValAspGluGlu 378

Qy      1081 GAAGATGATGAAGACTTAGACGAATTT-------------------------------- 1107
Db       379 GluIleAsp---ProLeuAspAlaPheMetAsnThrMetValLeuProGluValGluLys 397

Qy      1108 ---------------AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
Db       398 PheCysAsnGlyAlaProProProAlaValAsnAspGlyThrLeuAspSerLysMetAsn 417

Qy      1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAA------AATGAAAATTTAGATGAT 1206
Db       418 Gly------LysGluSerGlyAspArgProLysLysGlyPheAsnLysAlaLeuGlyArg 435

Qy      1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGA 1266
Db       436 IleIleGlnGly---GluAspSerAspSerAspTyrSerGlu------------------ 448

Qy      1267 AAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGAT 1326
Db       449 ---------ProLysAsnAspAspAspProSerLeuAspGluAspAspGluGluPheMet 465

Qy      1327 GAGCATATTAAAAAATATAAAAATGATAAG---------------CAGGTTAATAAGGAA 1371
Db       466 LysArgValLysLysThrLysAlaGluLysLeuSerLeuValAspHisSerLysIleGlu 485

Qy      1372 AAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAG 1431
Db       486 TyrGluProPheArgLysAsn---PheTyrIle---------------GluValLysAsp 499

Qy      1432 ATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
Db       500 IleSerArgMetThrGlnGluGluValAsnThrTyrArgLysGluLeu 515
```

RESULT 12
I38055
myosin heavy chain, perinatal skeletal muscle - human
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jul-2004
C;Accession: I38055; JH0154; S12459; S09332; A30220; S49478
R;Jullian, E.H.; Kelly, A.M.; Pompidou, A.J.; Hoffman, R.; Schiaffino, S.; Stedman, H.H
Eur. J. Biochem. 230, 1001-1006, 1995
A;Title: Characterization of a human perinatal myosin heavy-chain transcript.
A;Reference number: I38055; MUID:95324556; PMID:7601129
A;Accession: I38055
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1937 <RES>
A;Cross-references: UNIPROT:P13535; UNIPARC:UPI000016ADA9; EMBL:Z38133; NID:g558668; PI
R;Karsch-Mizrachi, I.; Feghali, R.; Shows, T.B.; Leinwand, L.A.
Gene 89, 289-294, 1990
A;Title: Generation of a full-length human perinatal myosin heavy-chain-encoding cDNA.
A;Reference number: JH0154; MUID:90323631; PMID:2373371
A;Accession: JH0154
A;Molecule type: mRNA
A;Residues: 1-14,'A',16-859 <KAR>
A;Cross-references: UNIPARC:UPI000017760F; GB:Y00821
A;Experimental source: skeletal muscle
R;Bober, E.
submitted to the EMBL Data Library, January 1989
A;Reference number: S12458
A;Accession: S12459
A;Molecule type: mRNA
A;Residues: 502-1071,'N',1073-1250,'DGG',1253-1376,'NT',1379-1913,'D',1915-1937 <BOB>
A;Cross-references: UNIPARC:UPI000016A5F8; EMBL:X51592; NID:g29465; PIDN:CAA35941.1; PI
A;Experimental source: clone gtMHC-F
R;Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
Eur. J. Biochem. 189, 55-65, 1990
A;Title: Identification of three developmentally controlled isoforms of human myosin hea
A;Reference number: S09331; MUID:90235862; PMID:1691980
A;Accession: S09332

```
A;Molecule type: mRNA
A;Residues: 502-547,'X',549-617,'X',619-687,'X',689-757,'X',759-827,'X',829-897,'X',899-
-1376,'NT',1379-1386,'X',1388-1456,'X',1458-1526,'X',1528-1596,'X',1598-1666,'X',1668-17
A;Cross-references: UNIPARC:UPI0000177610; EMBL:X51592
R;Feghali, R.; Leinwand, L.A.
J. Cell Biol. 108, 1791-1797, 1989
A;Title: Molecular genetic characterization of a developmentally regulated human perinat
A;Reference number: A30220; MUID:89234168; PMID:2715179
A;Accession: A30220
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 860-969,'Q',971-1246,'H',1248-1260,'G',1262-1296,'Q',1298-1503,'AH',1506-184
A;Cross-references: UNIPARC:UPI0000177611; GB:Y00821; NID:g34863; PIDN:CAA68757.1; PID:g
C;Genetics:
A;Gene: GDB:MYH8
A;Cross-references: GDB:125267; OMIM:160741
A;Map position: 17pter-17p12
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; nucleotide b
F;91-769/Domain: myosin motor domain homology <MMOT>
F;181-188/Region: nucleotide-binding motif A (P-loop)
F;551-588/Region: actin binding #status predicted
F;658-680/Region: actin binding #status predicted
F;842-1282/Region: S2 #status predicted
F;698,708/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                5.52e-10        Length:          1937
Score:                    317.50          Matches:         156
Percent Similarity:       42.08%          Conservative:    131
Best Local Similarity:    22.87%          Mismatches:      184
Query Match:              12.77%          Indels:          211
DB:                       2               Gaps:            35

US-09-837-344-45 (1-1482) x I38055 (1-1937)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||::::::   ::::::      |||      |||||::::   ||||||:::::::::
Db        930 GluGluGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                ||||:::|||::::::|||                          ::::::   |||
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||:::|||||||||      |||||||||      ||||||:::
Db        990 GluThrIleAlaLysLeuSerLysGluLysLysAlaLeuGlnGluThrHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                  ||||||:::   |||                        :::||||||   |||
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnIleLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||          ||||||||||||:::
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1049

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||   ||||||                     ||||||
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        268 AGCGATTTAGAACAAGAT-----------------AGACTTGCTAAAGAAAAGTTACAA 309
              |||:::|||   |||                 :::|||   |||::::::   :::
Db       1070 MetAspMetGluAsnAspLysGlnGlnLeuAspGluLysLeuGluLysLysGluPheGlu 1089

Qy        310 ----------------------------GAGCAGCAAAGCGATTTAGAACAAGAGAGA 339
                                         ||||||      :::::::::   ::::::
Db       1090 IleSerAsnLeuIleSerLysIleGluAspGluGlnAlaValGluIleGlnLeuGlnLys 1109

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||          :::::::::|||         ::::::|||   ||||||
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGlyGluGluIleGluAlaGluArg 1129

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   ::::::|||:::|||||||||   :::|||
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        448 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 495
                 |||:::|||:::|||              :::   :::|||      ::::::
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnValGluLeuAsnLysLys 1169

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 540
                 :::   :::   |||:::   :::   |||||||||:::
Db       1170 ArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        540 ------------------------------------------------------------ 540

Db       1190 AlaMetValAlaAlaLeuArgLysLysHisAlaAspSerMetAlaGluLeuGlyGluGln 1209

Qy        541 ------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
                         :::|||   |||:::||||||:::::::::::|||:::|||:::
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        589 GAG--------------------------------AGACGTGCTAAAGAAAAGTTGCAA 615
              |||                                :::||||||      |||:::
Db       1230 GluThrAspAspLeuSerSerAsnAlaGluAlaIleSerLysAlaLysGlyAsnLeuGlu 1249

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAA 675
              :::      |||   ||||||          |||            |||:::   |||
Db       1250 LysMetCysArgSerLeuGlu-----------AspGlnValSerGluLeuLysThrLys 1265

Qy        676 AAGGAACATGGAGATATATTAGCAGAGGATTTA---------TATGGTCGTTTAGAAATA 726
              :::|||           |||      ||||||           ||||||:::
Db       1266 GluGluGluGlnGlnArgLeuIleAsnAspLeuThrAlaGlnArgAlaArgLeuGlnThr 1285

Qy        727 CCAGCTATAGAACTTCCATCAGAA---AATGAACGTGGATATTATATACCACATCAATCT 783
              |||   |||         :::   :::|||:::         :::         |||
Db       1286 GluAlaGlyGluTyrSerArgGlnLeuAspGluLysAspAlaLeuVal---------Ser 1302

Qy        784 TCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATG------TCTATA 837
              |||   :::      :::  :::::::::      :::|||:::          :::
Db       1303 GlnLeuSerArgSerLysGlnAlaSerThrGlnGlnIleGluGluLeuLysHisGlnLeu 1322

Qy        838 ATAGAAAAAACAAAT---AGAGAATCTATTACAACAAATGTTGAAGGACGAAGG------ 888
              |||:::|||     :::   ::::::      ::::::      |||
Db       1323 GluGluGluThrLysAlaLysAsnAlaLeuAlaHisAlaLeuGlnSerSerArgHisAsp 1342

Qy        889 ---GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA------------ 933
                 |||:::   :::         ||||||:::::::::|||
Db       1343 CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluGlyLysAlaGluLeuGlnArg 1362

Qy        933 ------------------------------------------------------------ 933

Db       1363 AlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThrAspAla 1382

Qy        934 ---------------AAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCAT 978
                          :::   |||:::   :::   |||   |||
Db       1383 IleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeuGlnGlu 1402

Qy        979 ACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT 1038
                |||   |||         :::                 :::   |||   :::
Db       1403 AlaGluGluHisValGluAlaValAsnAlaLysCysAlaSerLeuGluLysThrLysGln 1422

Qy       1039 GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT------------ 1086
                 :::      |||   :::|||   ::::::|||   |||
Db       1423 ArgLeuGlnAsnGluValGluAspLeuMetLeuAspValGluArgSerAsnAlaAlaCys 1442

Qy       1087 ----------------------------GATGAAGACTTAGACGAATTTAAGCCTATT 1116
                                         |||:::   |||   |||:::|||
Db       1443 AlaAlaLeuAspLysLysGlnArgAsnPheAspLysValLeuSerGluTrpLysGln--- 1461

Qy       1117 GTGCAATATGACAATTTCCAAGATGAAGAAAACATA---------------------- 1152
                 :::|||:::      |||   |||
```

```
Db       1462 ---LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGlnLysGluSerArgSerLeu 1480

Qy       1153 ----------------------GGAATTTATAAAGAA-----------CTAGAAGAT 1176
                                    :::|||:::|||           ||||||
Db       1481 SerThrGluLeuPheLysValLysAsnValTyrGluGluSerLeuAspGlnLeuGluThr 1500

Qy       1177 TTGATAGAGAAAAATGAAAATTTA-----------GATGATTTAGATGAAGGAATAGAA 1224
              |||      :::|||:::||||||               ||||||   |||   |||
Db       1501 LeuArgArgGluAsnLysAsnLeuGlnGlnGluIleSerAspLeuThrGluGlnIleAla 1520

Qy       1225 AAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAAGGAAAGAAATATGAAAAA 1281
              :::      :::::::::   |||   ||||||||||||||     :::   ||||||
Db       1521 GluGlyGlyLysGlnIleHisGluLeuGluLysIleLysLysGlnValGluGlnGluLys 1540

Qy       1282 ACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAA 1338
                 :::          :::      :::   ||||||      |||      |||
Db       1541 CysGluIleGlnAlaAlaLeuGluGluAlaGluAlaSerLeuGluHisGluGluGlyLys 1560

Qy       1339 AAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTT 1398
              :::        ::::::|||:::   |||   :::      :::
Db       1561 IleLeuArgIleGlnLeuGluLeuAsnGlnValLysSerGluValAspArgLys------ 1578

Qy       1399 CATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
              |||   :::   |||  ||||||          |||:::|||      :::
Db       1579 ---IleAlaGluLysAspGluGluIle-----------AspGlnLeuLysArgAsnHis 1593

Qy       1459 ACTAAA 1464
              |||:::
Db       1594 ThrArg 1595

RESULT 13
A61231
myosin heavy chain nonmuscle form A - human
N;Alternate names: cellular myosin heavy chain; myosin type 9; NMMHC-A
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 14-Jul-1994 #text_change 31-Dec-2004
C;Accession: A61231; A34876; I52562; I61692
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on differe
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: A61231
A;Molecule type: mRNA
A;Residues: 1-715 <SIM>
A;Cross-references: UNIPROT:P35579; UNIPROT:Q9UMJ0; UNIPARC:UPI000016ADA3; GB:M69180; NI
R;Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A;Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through altern
A;Reference number: A34876; MUID:90138958; PMID:1967836
A;Accession: A34876
A;Molecule type: mRNA
A;Residues: 715-1961 <SAE>
A;Cross-references: UNIPARC:UPI000016ADA7; GB:M31013; NID:g189035; PIDN:AAA36349.1; PID:
R;Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.;
Blood 78, 1826-1833, 1991
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, c
A;Reference number: I52562; MUID:92003925; PMID:1912569
A;Accession: I52562
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-52,'EAI',56-659,'T',661-868,'T',870-930,'C',932-1239,'KG',1242-1337 <RES>
A;Cross-references: UNIPARC:UPI000016AD9A; GB:M81105; NID:g188988; PIDN:AAA59888.1; PID:
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A;Title: Identification and overlapping expression of multiple unconventional myosin ger
A;Reference number: A55758; MUID:94294418; PMID:8022818
A;Accession: I61692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 182-218 <BEM>
A;Cross-references: UNIPARC:UPI0000073258; GB:L29141; NID:g457249; PIDN:AAA20904.1; PID:
C;Genetics:
A;Gene: GDB:MYH9
A;Cross-references: GDB:120216; OMIM:160775
A;Map position: 22q12.3-22q13.1
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleoti
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1938/Domain: coiled coil #status predicted <COI>
F;837-1277/Domain: S2 #status predicted <DS2>
F;1278-1961/Domain: light meromyosin #status predicted <LMM>
F;1939-1961/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694,704/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                5.5e-10        Length:          1961
Score:                    317.50         Matches:         145
Percent Similarity:       43.98%         Conservative:    96
Best Local Similarity:    26.46%         Mismatches:      174
Query Match:              12.77%         Indels:          133
DB:                       1              Gaps:            26

US-09-837-344-45 (1-1482) x A61231 (1-1961)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   ::     ||||||:::|||     |||:::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAA---GAAAAGTTACAAGAGCAGCAAAGC 117
              :::          |||           ||||||   :::|||||:::|||:::|||
Db        966 LysValThrThrGlu-----------AlaLysLeuLysLysLeuGluGluGluGlnIle 981

Qy        118 GATTTAGAACAAGAG------AGACTTGCTAAAGAAAAG---------------------- 150
              ||||||   :::      :::|||||||||||||||
Db        982 IleLeuGluAspGlnAsnCysLysLeuAlaLysGluLysLysLeuLeuGluAspArgIle 1001

Qy        151 -----------------TTGCAAGAACAACAA-------------------------- 165
                              |||   |||::::::
Db       1002 AlaGluPheThrThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeu 1021

Qy        166 ----------------------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAG 201
                                   :::||||||||||:::      |||   :::||||||
Db       1022 LysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys 1041

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 261
                 |||:::   :::||||||:::   ||||||          ||||||:::
Db       1042 ---------GlnArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGly 1055

Qy        262 CAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 321
                 :::||||||      |||:::::|||      :::||||||   |||   :::
Db       1056 AspSerThrAspLeu---SerAspGlnIleAla------GluLeuGlnAlaGlnIleAla 1072

Qy        322 GATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 378
              :::|||:::   |||       :::   :::|||:::||||||          :::   :::
Db       1073 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgVal 1092

Qy        379 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 405
              |||:::|||                                |||      :::   :::
Db       1093 GluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGln 1112

Qy        406 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::   |||   |||   |||||||||   ||||||   ::::::   |||   ::::::
Db       1113 IleSerGluLeuGlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLys 1132

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG------------ 510
              |||:::   ||||||   :::|||      |||   |||   :::|||
Db       1133 GlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeu 1152

Qy        511 -----------------CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 552
                              ||||||   :::|||      |||||||||
```

```
Db       1153 AspSerThrAlaAlaGlnGlnGluLeuArgSerLysArgGluGlnGluValAsnIleLeu 1172

Qy        553 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA-----------CAAGAGAGACGTGCT 600
              :::|||      |||:::::::::                   ||||||   |||
Db       1173 LysLysThrLeuGluGluGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGln 1192

Qy        601 AAA------GAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||      :::   ::::::|||             |||||||||   |||
Db       1193 LysHisSerGlnAlaValGluGluLeuAlaGluGlnLeuGluGlnThrLysArg---Val 1211

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              |||   |||||||||:::   |||:::
Db       1212 LysAlaAsnLeuGluLysAlaLysGln-------------------------------- 1220

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
                                              ||||||||||||||
Db       1221 ----------------------------ThrLeuGluAsnGluArgGlyGluLeuAlaAsn 1231

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG------GAA 828
                      |||   |||      |||   :::             |||      |||
Db       1232 GluValLysValLeuLeuGlnGlyGlyArgAspSerGluHisLysArgLysLysValGlu 1251

Qy        829 ATGTCTATAATAGAA--------------------------------AAAACAAATAGA 855
              :::   |||                                    :::|||
Db       1252 AlaGlnLeuGlnGluLeuGlnValLysPheAsnGluGlyGluArgValArgThrGluLeu 1271

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
                    |||   ::::::        |||         |||   |||   :::
Db       1272 AlaAspLysValThrLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuSerGln 1291

Qy        916 AAG------------------------AAAGATGGTTCAATAAAACCAGAACAAAAAGAA 951
                                         ||||||   |||             |||   :::
Db       1292 SerAspSerLysSerSerLysLeuThrLysAspPheSerAlaLeuGluSerGlnLeuGln 1311

Qy        952 GATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAAT 1011
              |||           :::|||           :::   :::::::::|||
Db       1312 AspThrGlnGluLeuLeuGlnGluGluAsnArgGlnLysLeuSerLeuSerThr------ 1329

Qy       1012 GATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATA 1071
                ::::::   :::   |||||||||   ::::::   :::   :::   |||
Db       1330 ------LysLeuLysGlnValGluAspGluLysAsnSer---PheArgGluGlnLeuGlu 1346

Qy       1072 GATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAAT 1131
              :::|||||||||:::   :::         |||   :::             :::
Db       1347 GluGluGluGluGluAlaLysHisAsnLeuGluLysGlnIleAlaThrLeuHisAlaGln 1366

Qy       1132 TTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAAT 1191
                |||               :::   |||::::::||||||   :::
Db       1367 ValAlaAsp--------------MetLysLysLysMetGluAspSerValGlyCysLeu 1381

Qy       1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT------ 1245
              |||      :::::::::         ::::::|||      |||   ||||||
Db       1382 GluThrAlaGluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArg 1401

Qy       1246 ---GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAA 1302
                 |||||||||:::               |||   ||||||||||||             :::
Db       1402 HisGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeuGlnGln 1421

Qy       1303 CCAAATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                |||      |||   |||   :::|||      :::   ::::::
Db       1422 GluLeuAspAspLeuLeuValAspLeuAspHis------GlnArgGlnSerAlaCysAsn 1439

Qy       1360 GTTAATAAGGAAAAGGAAAAATTC 1383
              :::   |||:::::::::||||||
Db       1440 LeuGluLysLysGlnLysLysPhe 1447

RESULT 14
A47297
myosin heavy chain form B, nonmuscle - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A47297; A55441
R;Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.
Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993
A;Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscl
A;Reference number: A47297; MUID:93219383; PMID:8464900
A;Accession: A47297
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1992 <BHA>
A;Cross-references: UNIPROT:Q04834; UNIPARC:UPI00000FC054; GB:L09740; NID:g214623; PIDN:
A;Experimental source: XTC cells
A;Note: sequence extracted from NCBI backbone (NCBIP:128722)
R;Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.
J. Biol. Chem. 270, 1395-1401, 1995
A;Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34(
A;Reference number: A55441; MUID:95138137; PMID:7836406
A;Accession: A55441
A;Status: preliminary
A;Molecule type: protein
A;Residues: 198-232 <KEL>
A;Cross-references: UNIPARC:UPI0000177615
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-787/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.:              5.49e-10        Length:         1992
Score:                  317.50          Matches:        133
Percent Similarity:     45.94%          Conservative:   116
Best Local Similarity:  24.54%          Mismatches:     171
Query Match:            12.77%          Indels:         122
DB:                     2               Gaps:           23

US-09-837-344-45 (1-1482) x A47297 (1-1992)

Qy           1 CAAGAACAACAAAGCGATCTAGAACAA---------GAGAGACGTGCTAAAGAAAAGTTG 51
               |||             |||||||||:::         |||      |||   :::||||||
Db         965 GlnThrHisValGlnAspLeuGluGluGlnLeuAspGluGluGluAlaAlaGlnLysLeu 984

Qy          52 CAAGAACAACAA---------------------AGCGATTTAGAACAAGATAGACTTGCT 90
               |||   ::::::                        :::|||:::|||   |||
Db         985 GlnLeuGluLysValThrAlaGluAlaLysIleLysLysMetGluGluAspIleLeuVal 1004

Qy          91 AAAGAAAAGTTACAAGAGCAGCAAAGCGAT---TTAGAACAAGAGAGACTTGCTAAAGAA 147
                        |||::::::|||   |||      |||:::::::::::|||   :::|||
Db        1005 ---------LeuGluAspGlnAsnSerLysPheLeuLysGluLysLysLeuLeuGluGlu 1021

Qy         148 AAGTTGCAAGAACAACAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAA------ 198
               ::::::   |||       |||   |||   |||:::|||   :::||||||
Db        1022 ArgIleAlaGluSerThrSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAla 1041

Qy         199 AAGTTGCAAGAACAACAA------------AGCGATTTAGAACAAGAG------------ 234
               ||||||:::   :::|||             ||||||||||||:::
Db        1042 LysLeuLysAsnLysGlnGluMetMetIleSerAspLeuGluGluArgLeuLysLysGlu 1061

Qy         235 ---------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273
                                    :::|||||   ||||||      :::   :::|||
Db        1062 GluLysThrArgGlnGluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAsp 1081

Qy         274 TTAGAACAAGATAGACTTGCTAAA--------------GAAAAGTTACAAGAGCAGCAA 318
               ||||||::::::|||:::                  |||:::|||:::   |||
Db        1082 Phe---GlnAspGlnIleAlaGluLeuGlnAlaGlnIleGluGluLeuLysLeuGlnLeu 1100

Qy         319 AGCGATTTAGAACAAGAG--------------AGACGTGCTAAAGAAAAGTTGCAAGAA 363
               :::      |||:::|||                  |||      |||::::::   :::
Db        1101 AlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgGlyAspGluGluValLeuGln 1120

Qy         364 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 423
               :::   :::   |||:::      |||      :::   ::::::   |||   |||
Db        1121 LysAsnAsnThrLeuLysLeuValArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGlu 1140

Qy         424 GATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 483
               |||||||||   |||:::   ::::::   |||   ::::::|||:::   ||||||
Db        1141 AspLeuGluSerGluLysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSer 1160
```

```
Qy        484 CAAGAGAGACGTGCT---------------------------------------------- 498
              :::|||      |||
Db       1161 GluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGln 1180

Qy        499 --------------------AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 537
                                 ::::::::::::   |||   :::      :::|||:::
Db       1181 GlnGluLeuArgThrLysArgGluGlnGluValAlaGluLeuArgLysSerIleGluGlu 1200

Qy        538 GAGAGACGTGCTAAAGAA----------------------------------AAGTTG 561
              |||   |||      |||                                      |||
Db       1201 GluThrArgAsnHisGluAlaGlnIleGlnGluMetArgGlnArgGlnAlaThrAlaLeu 1220

Qy        562 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAG 621
              :::|||            |||||||||   :::|||   |||      |||::::::
Db       1221 GluGluLeuSerGluGlnLeuGluGlnAlaLysArgPheLysValAsnLeuGluLysAsn 1240

Qy        622 CAAAGAGATTTAGAA---------CAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
              ::::::   ||||||         :::      |||         |||:::|||::::::
Db       1241 LysGlnSerLeuGluSerAspAsnLysGluLeuAlaThrGluValLysSerLeuGlnGln 1260

Qy        673 AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA---GAAATACCA 729
                 |||         :::         :::   |||   |||::::::   |||:::
Db       1261 MetLysAlaGluSerGluTyrLysArgLysLysLeuGluGlyGlnValGlnGluLeuHis 1280

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              |||      |||      :::      |||         :::         |||   |||
Db       1281 AlaLysValLeuGluGlyAspArgLeuArgAlaAspMetValGluLysSerSerLysLeu 1300

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                  :::          |||                  |||::::::|||:::
Db       1301 GlnAsnGluLeuGluAsnValSer----------------SerLeuLeuGluGluAla 1314

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
                 ::::::   |||
Db       1315 GluLysLysGlyIle---------------------------------------------- 1319

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              :::      ::::::   :::         |||   :::|||                :::
Db       1320 LysLeuAlaLysAspValAlaSerMetGluSerGlnLeuGlnAspThrGlnGluLeuLeu 1339

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
              |||      |||   :::   :::|||  |||                 :::|||   :::
Db       1340 GlnGluGluThrArgGlnLysLeuAsnGlnSerSer------------ArgIleArgGln 1355

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                 |||:::|||                :::::::|||   :::::::::|||::::::
Db       1356 LeuGluGluGlu--------------LysAsnAsnLeuGlnGluGlnGlnGluGluGlu 1370

Qy       1090 GAAGACTTAGACGAATTT------AAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
              |||:::      :::         |||   |||:::         :::
Db       1371 GluGluAlaArgLysSerLeuGluLysGlnIleLeuSerLeuGlnSer------------ 1386

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203
              :::   |||          |||::::::::::|||   :::          |||   |||:::
Db       1387 GlnLeuIleGluAlaLysLysLysValAspAspGluValGlyThrIleGluGlyLeuGlu 1406

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAAAAA 1254
              ::::::   :::   :::   |||   :::|||   |||                |||||||||
Db       1407 GluValLysLysLysLeuLeuLysAspThrGluGlyLeuGlyGlnArgLeuGluGluLys 1426

Qy       1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
              |||          :::|||  ||||||||||||   |||   |||   |||   :::
Db       1427 IleIleAlaTyrGluLysLeuGluLysThrLys---AsnArg-LeuGlnGlnGluLeuAs 1445

Qy       1315 AGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
                    |||:::   :::            :::::::::|||   |||   :::|||::
Db       1445 pAspLeuMetValAspLeuAspHisGlnArgGlnIleValSerAsnLeuGluLysLysGl 1465

Qy       1375 GAAA 1378
              :|||
Db       1465 nLys 1466
```

RESULT 15
B85431
trichohyalin like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: B85431
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spri
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: B85431
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1432 <STO>
A;Cross-references: UNIPROT:O23230; UNIPARC:UPI000009D2F0; GB:NC_001268; NID:g7270600;
C;Genetics:
A;Gene: AT4g36520
A;Map position: 4

Alignment Scores:

Pred. No.:	6.16e-10	Length:	1432
Score:	317.00	Matches:	136
Percent Similarity:	43.26%	Conservative:	108
Best Local Similarity:	24.11%	Mismatches:	196
Query Match:	12.75%	Indels:	124
DB:	2	Gaps:	22

US-09-837-344-45 (1-1482) x B85431 (1-1432)

```
Qy         19 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 78
              :::|||:::|||   |||   ||||||            :::::::::|||      :::
Db        632 MetGluGluGluThrArgIleLysGluAlaArgLeuArgGluGluAsnAsp---ArgArg 650

Qy         79 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAG---------CAAAGCGATTTAGAACAA 129
              :::|||:::|||   ||||||   :::   :::         ::::::   |||||||||
Db        651 GluArgValAlaValGluLysAlaGluAsnGluLysArgLeuLysAlaAlaLeuGluGln 670

Qy        130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189
              |||          :::   |||:::::::|||   :::         |||   |||||||||
Db        671 GluGlu---LysGluArgLysIleLysGluAlaArgGluLysAlaGluAsnGluArgArg 689

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
              |||   |||      ::::::                |||||||||||:::   ||||||
Db        690 AlaValGluAlaArgGluLys-----------AlaGluGlnGluArgLysMetLysGlu 705

Qy        250 -----------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTT 291
                              :::|||:::|||          |||::::::|||
Db        706 GlnGlnGluLeuGluLeuGlnLeuLysGluAlaPheGluLysGluGluGluAsnArgArg 725

Qy        292 GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 351
                 :::|||          :::|||         |||:::|||||||||   ||||||
Db        726 MetArgGluAlaPheAlaLeuGluGln---------GluLysGluArgArgIleLysGlu 742

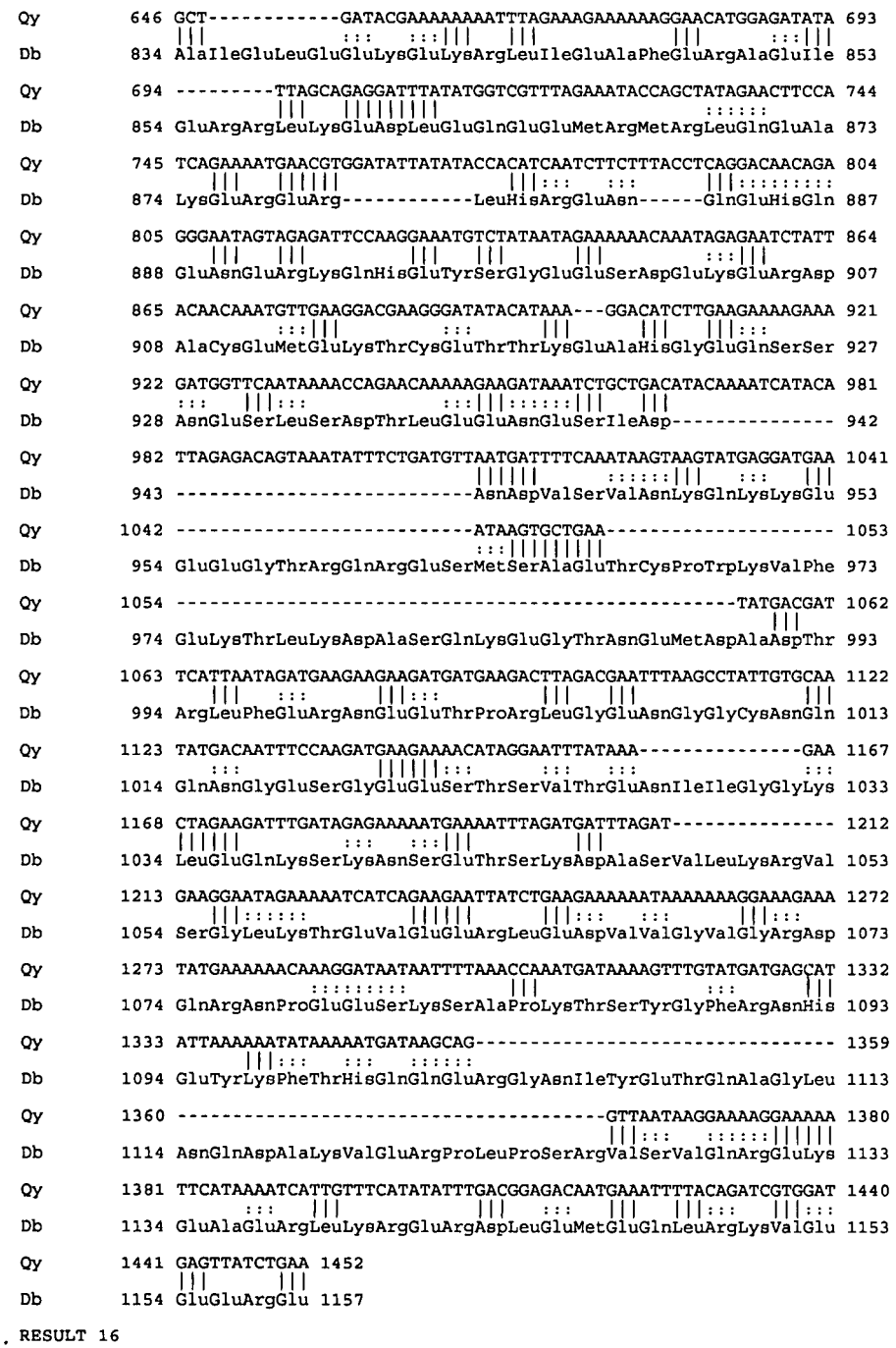
Qy        352 AAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGACGTGCTAAAGAAAAGTTG 408
                    :::::::::::::::::         :::|||   |||   |||   :::|||
Db        743 AlaArgGluLysGluGluAsnGluArgArgIleLysGluAlaArgGluLysAlaGluLeu 762

Qy        409 CAAGAACAA---CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAA 465
              :::::::::   ::::::   ||||||||||||          :::   :::::::::|||
Db        763 GluGlnArgLeuLysAlaThrLeuGluGlnGluGlu---LysGluArgGlnIleLysGlu 781

Qy        466 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 525
              :::|||          |||   ||||||||||||||||   |||::::::
Db        782 ArgGlnGluArgGluGluAsnGluArgArgAlaLysGluValLeuGluGln--------- 798

Qy        526 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 585
                    |||   ||||||:::   ||||||   |||:::::::::            |||
Db        799 ---AlaGluAsnGluArgLysLeuLysGluAlaLeuGluGlnLys------------Glu 813

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAG 645
                 |||||||||   ||||||      ::::::::::::         :::   |||:::
Db        814 AsnGluArgArgLeuLysGluThrArgGluLysGluGluAsnLysLysLysLeuArgGlu 833
```



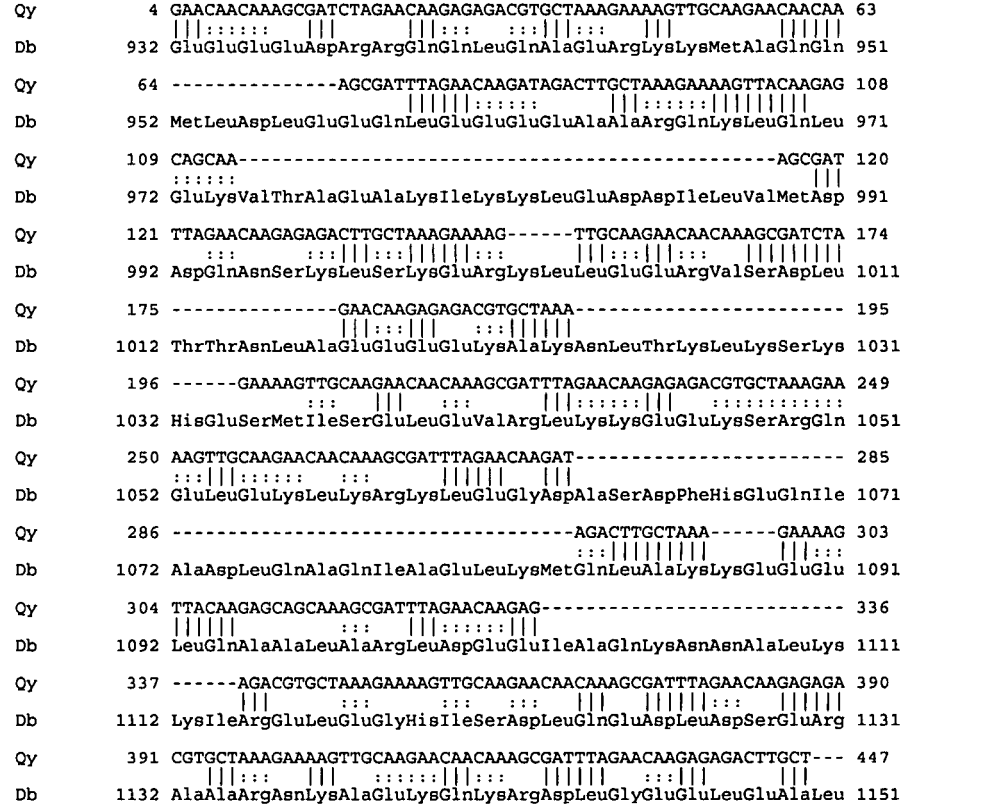
RESULT 16

JC5421
smooth muscle myosin heavy chain 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C;Accession: JC5421
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5421
A;Molecule type: mRNA
A;Residues: 1-1938 <HAS>
A;Cross-references: UNIPROT:O08638; UNIPARC:UPI000002A11C; DDBJ:D85924; NID:g1945079; PI
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:

Pred. No.:	5.86e-10	Length:	1938
Score:	317.00	Matches:	144
Percent Similarity:	42.02%	Conservative:	114
Best Local Similarity:	23.45%	Mismatches:	180
Query Match:	12.75%	Indels:	176
DB:	2	Gaps:	27

US-09-837-344-45 (1-1482) x JC5421 (1-1938)



```
Qy       448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
Db      1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy       502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
Db      1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy       535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
Db      1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy       580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db      1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy       640 AGGAAGGCTGAT----------------------------ACGAAAAAAAATTTAGAA 669
Db      1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy       670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db      1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu---------------------- 1264

Qy       730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
Db      1265 ----------------------------------------------GlnSerLysCys 1268

Qy       790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
Db      1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy       850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db      1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy       901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
Db      1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy       961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
Db      1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy      1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db      1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy      1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db      1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy      1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db      1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy      1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
Db      1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy      1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT---------------------- 1290
Db      1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy      1291 ------------------------------------AATAATTTTAAACCAAATGAT 1311
Db      1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy      1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT---------------------- 1350
Db      1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy      1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
Db      1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy      1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
Db      1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502
```

RESULT 17
JC5420
smooth muscle myosin heavy chain 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C;Accession: JC5420
R;Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A;Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A;Reference number: JC5420; MUID:97242182; PMID:9125171
A;Accession: JC5420
A;Molecule type: mRNA
A;Residues: 1-1972 <HAS>
A;Cross-references: UNIPROT:O08638; UNIPARC:UPI00000279BF; DDBJ:D85923; NID:g1945077; P
A;Experimental source: smooth muscle
C;Comment: This protein plays a role in smooth muscle cell contraction.
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:

Pred. No.:	5.84e-10	Length:	1972
Score:	317.00	Matches:	144
Percent Similarity:	42.02%	Conservative:	114
Best Local Similarity:	23.45%	Mismatches:	180
Query Match:	12.75%	Indels:	176
DB:	2	Gaps:	27

US-09-837-344-45 (1-1482) x JC5420 (1-1972)

```
Qy         4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db       932 GluGluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy        64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
Db       952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy       109 CAGCAA-----------------------------------------------AGCGAT 120
Db       972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
Db       992 AspGlnAsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeu 1011

Qy       175 ---------------GAACAAGAGAGACGTGCTAAA---------------------- 195
Db      1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLys 1031

Qy       196 ------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249
Db      1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy       250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---------------------- 285
Db      1052 GluLeuGluLysLeuLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle 1071

Qy       286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
Db      1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy       304 TTACAAGAGCAGCAAAGCGATTTAGAACAAGAG---------------------- 336
```

```
Db       1092 LeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAlaLeuLys 1111

Qy        337 ------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
                  |||      :::      :::   :::   |||   ||||||:::   ||||||
Db       1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                 |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db       1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeu 1151

Qy        448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
                |||   :::|||::::::                :::      :::||||||   |||||||||
Db       1152 LysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy        502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                  :::                        |||   |||:::            |||
Db       1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy        535 ---------CAAGAGAGACGTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGAT 579
                      ||||||   |||   |||      :::   :::::::|||
Db       1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThrGluGln 1211

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                |||||||||   :::|||||||||       |||   :::   ::::::   ||||||:::
Db       1212 LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeuGluLys 1231

Qy        640 AGGAAGGCTGAT-----------------------------ACGAAAAAAAATTTAGAA 669
                  ||||||                                |||:::   :::|||
Db       1232 GluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGlu 1251

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
                |||||||:::      ::::::      :::||||||
Db       1252 HisLysLysLysLysLeuGluValGlnLeuGlnAspLeu---------------------- 1264

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
                                                                  ||||||
Db       1265 ------------------------------------------------GlnSerLysCys 1268

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
                     |||      :::      |||   ::::::      ::::::
Db       1269 SerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLysLeuGln---Asn 1287

Qy        850 AATAGAGAATCTATTACA---------ACAAATGTTGAAGGACGAAGGGATATACATAAA 900
                  ||||||:::|||                  ||||||:::
Db       1288 GluValGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAla------------ 1303

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
                  ::::::      ||||||   :::              |||   :::|||
Db       1304 ------IleLysLeuAlaLysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGln 1321

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
                  :::|||      |||   :::   :::|||:::|||               :::
Db       1322 GluLeuLeuGlnGluGluThrArgGlnLysLeuAsnValSerThr------------Lys 1337

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
                :::   :::   |||||||||   ::::::          |||   |||   :::|||
Db       1338 LeuArgGlnLeuGluAspGluArgAsnSer---LeuGlnAspGlnLeuAspGluGluMet 1356

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
                |||      ::::::|||:::                |||          |||   |||
Db       1357 GluAlaLysGlnAsnLeuGluArgHis---------ValSerThrLeuAsnIleGlnLeu 1373

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
                  ::::::           |||:::|||:::|||                |||   :::
Db       1374 SerAspSer---------LysLysLysLeuGlnAspPheAlaSerThrIleGluValMet 1390

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT---------GAAGAA 1251
                ::::::      :::   :::::::|||      |||   ||||||          ||||||
Db       1391 GluGluGlyLysLysArgLeuGlnLysGluMetGluGlyLeuSerGlnGlnTyrGluGlu 1410

Qy       1252 AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT-------------------- 1290
                |||                |||   |||||||||||:::
Db       1411 LysAlaAlaAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAsp 1430

Qy       1291 ------------------------------------AATAATTTTAAACCAAATGAT 1311
                                                    :::|||   :::
Db       1431 AspLeuValValAspLeuAspAsnGlnArgGlnLeuValSerAsnLeuGluLysLysGln 1450

Qy       1312 AAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT-------------------- 1350
                |||      :::|||:::   :::   :::   ||||||
Db       1451 LysLys---PheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAsp 1469

Qy       1351 ------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATA 1404
                      |||:::          :::   :::||||||      |||:::|||      :::
Db       1470 GluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu---SerLeu 1488

Qy       1405 TTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTA 1446
                        |||   |||:::      :::||||||
Db       1489 AlaArgAlaLeuGluGluAlaLeuGluAlaLysGluGluLeu 1502

RESULT 18
T14867
interaptin - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T14867
R;Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A;Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dictyo
ts.
A;Reference number: Z18248; MUID:98365468; PMID:9700162
A;Accession: T14867
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1738 <RIV>
A;Cross-references: UNIPROT:O76329; UNIPARC:UPI000007F02C; EMBL:AF057019; NID:g3549260;
C;Genetics:
A;Gene: abpD
A;Introns: 173/2; 1680/1

Alignment Scores:
Pred. No.:              6.34e-10        Length:         1738
Score:                  316.50          Matches:        127
Percent Similarity:     46.86%          Conservative:   127
Best Local Similarity:  23.43%          Mismatches:     171
Query Match:            12.73%          Indels:         117
DB:                     2               Gaps:           24

US-09-837-344-45 (1-1482) x T14867 (1-1738)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA--------------GAA 45
                |||:::   ||||||   |||   ::::::|||:::   :::              |||
Db       1012 GlnGlnLeuGlnSerGlnLeuAsnGluGlnArgGlnGlnGlnSerAsnGlnLeuSerGlu 1031

Qy         46 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 105
                |||   |||:::            :::|||:::::::::     :::::::::   |||
Db       1032 LysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGlnPheAspGlnLysGluGlnGln 1051

Qy        106 GAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAA---AAGTTGCAAGAACAA 162
                 :::|||      :::|||   :::        ||||||   ::::::|||:::
Db       1052 LeuLysGlnGlnSerIleGluAsnAspLeuPheGluLysGluAsnGlnIleGlnGlnLeu 1071

Qy        163 CAAAGCGATCTA---GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 219
                ||||||   |||   ||||||   ::::::         :::|||   |||:::
Db       1072 GlnSerGlnLeuAsnGluGlnArgGlnGlnGlnSerAsnGlnLeuSerGluLysAspGln 1091

Qy        220 GATTTAGAACAA--------------GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 264
                  |||   |||                |||          ||||||:::      :::|||
Db       1092 GlnLeuAsnGlnLeuIleGluLysAsnGluSerAspGlnLysGluGlnGlnLeuLysGln 1111

Qy        265 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA---AAGTTACAAGAGCAGCAAAGC 321
                ||||||   :::|||   |||   :::   ||||||   ::::::|||:::   |||
Db       1112 GlnSer---IleGluAsnAspLeuIleGluLysGluAsnGlnIleGlnGlnLeuGlnLeu 1130

Qy        322 GATTTAGAACAAGAGAGACGT---------------------------------------- 342
                  |||   ::::::|||:::
```

```
Db       1131 GlnLeuAsnGluGlnArgGlnLeuGlnSerGluValSerIleAspAsnAspLysIleLeu 1150

Qy        343 ---GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA-----------GAACAAGAG 387
              :::::::::|||:::::::  ||||||||||||              :::::::::
Db       1151 GluLeuGluLysGlnLeuLysGlnCysGlnSerAspLeuLeuLysLeuAsnAspGluLys 1170

Qy        388 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT 447
              ::::::      ::::::||||||:::::::|||   :::   :::|||   :::|||
Db       1171 GlnGlnGlnAspLysGlnLeuGlnAspLysGlnIleGluPheAspGlnLeuGlnLeuThr 1190

Qy        448 ----------------------------------------------------------AAA 450
                                                                         |||
Db       1191 PheAsnGlnPheLysAsnAspLysAspSerGlnPheIleGlnLeuGlnAspAspGlnLys 1210

Qy        451 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
              ::::::||||||      |||   ||||||   |||            |||::::::
Db       1211 GlnGlnLeuGlnSerIleGlnGlnAspLeuAsnGln---------LeuLysGlnGluAsn 1227

Qy        511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
              ||||||::::::      |||   ::::::            :::   :::|||   :::
Db       1228 GlnGluLysGluLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnPheGlu 1247

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAA---GAAAAGTTGCAAGAGCAGCAAAGA 627
                 :::   |||::::::          |||   ||||||||||||      |||:::
Db       1248 AsnGlnGluLysGluLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGln 1267

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAA---------AAAAATTTAGAAAGAAAAAAG 678
              :::|||   |||         :::   :::        |||            |||
Db       1268 AsnLeuAsnGlnLeuAsnAspGluAsnGlnGluLysValLysGlnPheSerGluLysAsp 1287

Qy        679 GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAA 738
              |||                :::   :::||||||                        :::
Db       1288 GluLysLeuGlnSerIleGlnGlnAspLeuAsn----------------------Gln 1299

Qy        739 CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGAC 798
              |||       ||||||:::                                    :::::
Db       1300 LeuLysGlnGluAsnGlnGlu--------------------------------LysGlu 1308

Qy        799 AACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---------ATAGAAAAAACA 849
                 :::   :::   :::|||   |||      ||||||         :::   :::
Db       1309 LysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAspLeuAsnGlnLeu 1328

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              |||   :::   |||       |||   :::                          |||
Db       1329 AsnAspAspGlnIleLysLysAsnGluLys--------------------------Leu 1339

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              :::||||||:::::      :::|||   :::|||      :::
Db       1340 LysGluLysGluGluGlnLeuLeuLysLeuGlnGlnAspPheAsnAspGlnGlnSerGln 1359

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
              |||       ||||||                        |||      |||      |||
Db       1360 GlnLeuLysGlnLeuGluGluLysLeuSerGluLysGluAsnGlnLeuGlnGlnLeuLys 1379

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
                 |||:::||||||                  :::   |||     :::::::::   :::
Db       1380 GlnGluAsnGluIle--------------AsnGlnLeuAsnGlnGlnGlnGlnSerAsn 1394

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
              |||   :::   :::   |||      :::                 |||:::::::::
Db       1395 GluIleIleGlnGlnLeuLysAspGlnLeuLeuLysGlnGlnGlnGlnGluGlnGlnGlu 1414

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
                    ||||||:::|||   ||||||::::::   |||   |||
Db       1415 AsnAsnAsnGluLysGluIleGluArgLeuIleGlnGluIleGluGlnLeuLysGlnGln 1434

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA--------- 1260
                 |||   |||::::::|||      |||||||||   ::::::||||||
Db       1435 GlnGlu---IleAspGlnSer------GluLeuSerAsnLysGluIleLysIleGlnThr 1451

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
                 :::::::::::::::          |||      |||
Db       1452 ThrGlnGlnGluPheAspGlnLeuSerHisAsnArgSerLys----------------- 1465
```

```
Qy       1321 TATGATGAG---CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
                 |||:::   |||:::::::::   :::      ::::::::::::::   |||:::
Db       1466 ---AspGlnLeuHisLeuGlnGlnLeuGln------GlnGluLeuAspGlnLeuLysGln 1482

Qy       1378 AAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTG 1437
                 |||                          |||   |||::::::      :::::::::
Db       1483 SerPhe--------------------AspAspGlnAspHisGlnPheLysLysValIle 1495

Qy       1438 GATGAG 1443
                 ||||||
Db       1496 AspGlu 1497
```

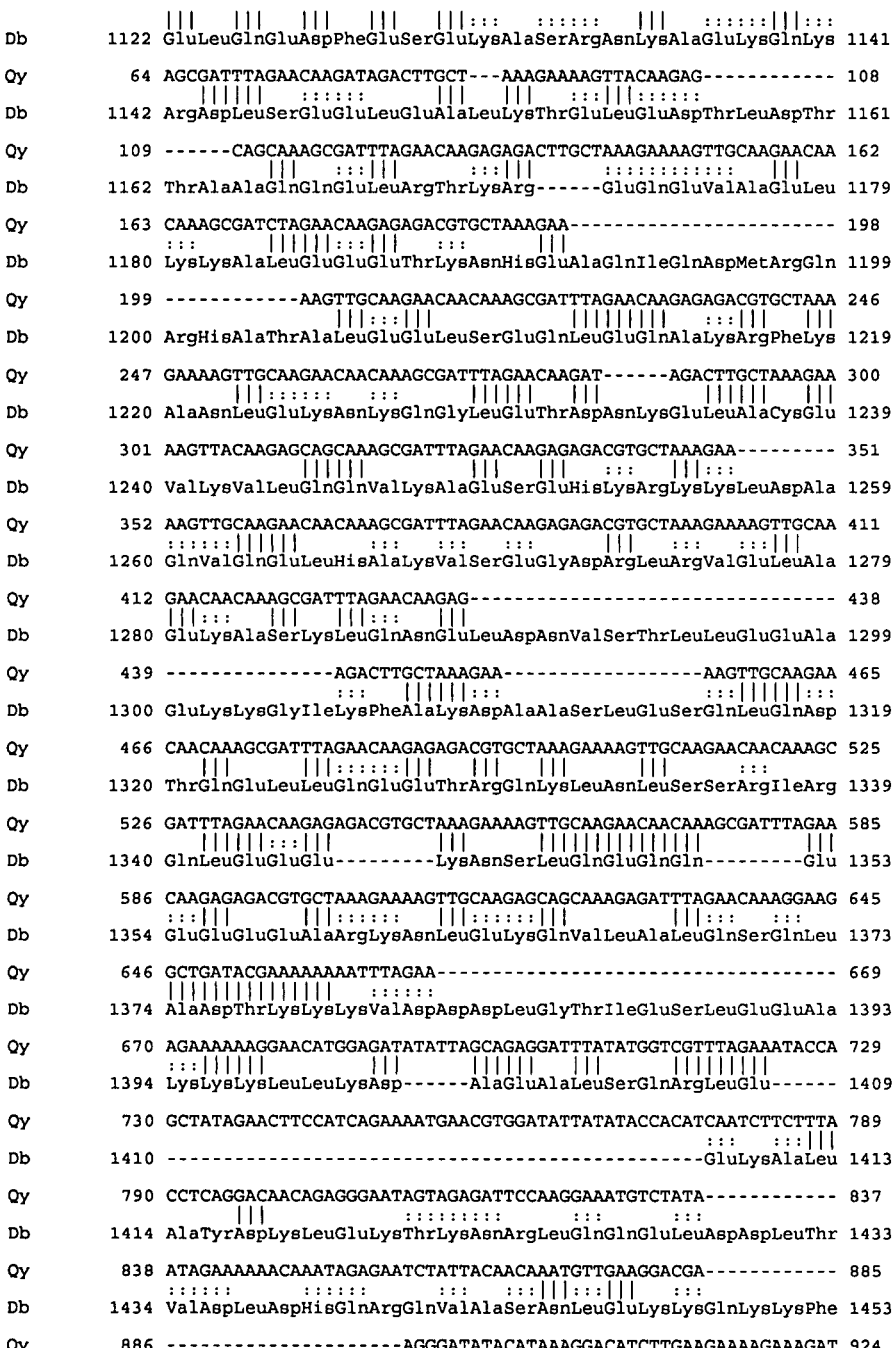
RESULT 19
A59252
myosin heavy chain, nonmuscle, form IIB - human
N;Alternate names: myosin type 10; NMMHC-B
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A59252; B61231; G02055
R;Phillips, C.L.; Yamakawa, K.; Adelstein, R.S.
J. Muscle Res. Cell. Motil. 16, 379-389, 1995
A;Title: Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis
A;Reference number: A59252; MUID:96025307; PMID:7499478
A;Accession: A59252
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1976 <SIM>
A;Cross-references: UNIPROT:P35580; UNIPARC:UPI000012FB83; GB:M69181; NID:g641957; PIDN
A;Experimental source: clone lib Lambda Zap II adult human T-cell library; cell line Jur
A;Note: between nucleotides 1942-1943 in mRNA encoding human brain MHC-B there is an alt
R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein,
Circ. Res. 69, 530-539, 1991
A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on differe
A;Reference number: A61231; MUID:91316803; PMID:1860190
A;Accession: B61231
A;Molecule type: mRNA
A;Residues: 63-237,'K',239-664,'L',666-722 <SI2>
A;Cross-references: UNIPARC:UPI000017761F; GB:M69181; NID:g641957
R;Weir, L.
submitted to the EMBL Data Library, August 1995
A;Reference number: H00753
A;Accession: G02055
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-81 <WEI>
A;Cross-references: UNIPARC:UPI000006EC6B; EMBL:U34304; NID:g1143217; PIDN:AAA84880.1; P
C;Genetics:
A;Gene: GDB:MYH10
A;Cross-references: GDB:127350; GDB:G00-127-350; OMIM:160776
A;Map position: 17p13-17p13
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F;88-771/Domain: myosin motor domain homology <MMO>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;559-572/Region: actin binding #status predicted
F;633-647/Region: actin binding #status predicted
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;701,711/Active site: Cys #status predicted

Alignment Scores:

Pred. No.:	6.59e-10	Length:	1976
Score:	316.00	Matches:	144
Percent Similarity:	40.48%	Conservative:	109
Best Local Similarity:	23.04%	Mismatches:	180
Query Match:	12.71%	Indels:	192
DB:	2	Gaps:	25

US-09-837-344-45 (1-1482) x A59252 (1-1976)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
```



RESULT 20
A59293
skeletal myosin heavy chain - domestic rabbit
C;Species: Oryctolagus cuniculus
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: A59293
R;Maeda, K.; Hostinova, E.; Roesc.Kleinkauf, A.; Schuster, H.; Gasperik, J.; Wittinghofe
submitted to GenBank, July 1995
A;Description: Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal mus
A;Reference number: A59293
A;Accession: A59293
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1938 <MAE>
A;Cross-references: UNIPROT:Q28641; UNIPARC:UPI000012FB71; GB:U32574; NID:g940232; PIDN:
A;Experimental source: strain New Zealand White; cell type skeletal muscle fiber type II
C;Genetics:
A;Gene: MHC
C;Superfamily: myosin heavy chain; myosin motor domain homology

F;89-769/Domain: myosin motor domain homology <MMO>

Alignment Scores:

Pred. No.:	8.42e-10	Length:	1938
Score:	314.00	Matches:	152
Percent Similarity:	42.32%	Conservative:	129
Best Local Similarity:	22.89%	Mismatches:	179
Query Match:	12.63%	Indels:	204
DB:	2	Gaps:	36

US-09-837-344-45 (1-1482) x A59293 (1-1938)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::      |||      ||||||:::   ||||||:::::::::
Db        930 GluAspGluGluGluIleAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 949

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                |||:::|||::::::|||                          ::::::   |||
Db        950 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 969

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        970 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 989

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::         :::|||   |||||||||      |||||||||      ||||||:::
Db        990 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1009

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                  |||||||:::   |||                      :::||||||   |||
Db       1010 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1029

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGACGTGCTAAAGAA 249
              |||:::::::|||       |||             ||||||||||||:::::::   :::
Db       1030 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysIleArgMet 1049

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
              |||:::       :::       ||||||   |||   :::|||||||:::|||
Db       1050 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerThr 1069

Qy        307 CAAGAGCAGCAAAGCGAT---------TTAGAACAAGAGAGACGTGCTAAAGAA------ 351
                 :::   :::::::|||         |||:::::::::   :::   ||||||
Db       1070 MetAspIleGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysGluPheGlu 1089

Qy        352 ------------------AAGTTGCAAGAACAACAAAGC---GATTTAGAACAAGAGAGA 390
                               |||::::::::::::|||:::       ::::::   ::::::
Db       1090 MetSerAsnLeuGlnSerLysIleGluAspGluGlnAlaLeuAlaMetGlnLeuGlnLys 1109

Qy        391 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
              :::   ||||||         :::::::::|||   :::   ::::::|||   ||||||
Db       1110 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1129

Qy        442 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACGTGCT 498
                 ::::::   |||   :::::::|||:::|||||||||   :::|||
Db       1130 AlaSerArgAlaLysAlaGluLysGlnArgSerAspLeuSerArgGluLeuGluGluIle 1149

Qy        499 AAAGAAAAGTTGCAAGAA------------CAACAAAGCGATTTAGAACAAGAGAGACGT 546
                 |||:::|||:::|||                    :::   :::|||      ::::::
Db       1150 SerGluArgLeuGluGluAlaGlyGlyAlaThrSerAlaGlnIleGluMetAsnLysLys 1169

Qy        547 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG-------------- 591
                 :::   :::   |||:::   :::   |||||||||:::
Db       1170 ArgGluAlaGluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGlnHisGlu 1189

Qy        591 ------------------------------------------------------------ 591

Db       1190 AlaThrAlaAlaThrLeuArgLysLysHisAlaAspSerValAlaGluLeuGlyGluGln 1209

Qy        592 ------------AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                         :::|||   |||:::||||||:::::::::::   :::|||:::
Db       1210 IleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeuLysMet 1229

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACATGGAGATATATTA 696

                    |||         |||:::|||         |||      |||::::::
Db       1230 GluIleAspAspLeuAlaSerAsnMetGluThrValSerLysAlaLysGlyAsnLeu--- 1248

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAA 756
                 |||   :::       ||||||         ||||||   ::::::   |||
Db       1249 ---GluLysMetCysArgThrLeuGluAspGlnValSerGluLeuLysThrLysGluGlu 1267

Qy        757 CGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGA 816
                                  ||||||      :::         :::         :::|||
Db       1268 Glu---------------HisGlnArgLeuIleAsnAspLeuSerAlaGlnArgAlaArg 1282

Qy        817 ---------GATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
                        :::|||   |||   |||      :::         ::::::|||:::
Db       1283 LeuGlnThrGluSerGlyGluPheSer---ArgGlnLeuAspGluLysAspSerLeuVal 1301

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
              :::   :::       :::                 :::||||||   |||
Db       1302 SerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLys------ 1319

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
                 :::      ::::::|||   ||:::                  |||   |||:::
Db       1320 ------ArgGlnLeuGluGluGluIleLysAlaLysSerAlaLeuAlaHisAlaLeuGln 1337

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
              :::                       |||             :::||||||:::|||
Db       1338 SerAlaArgHisAsp------CysAspLeuLeuArgGluGlnTyrGluGluGluGlnGlu 1355

Qy       1048 GCT------------------------------------------------------- 1050
              |||
Db       1356 AlaLysAlaGluLeuGlnArgAlaMetSerLysAlaAsnSerGluValAlaGlnTrpArg 1375

Qy       1051 ---GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                 :::|||:::         |||      |||:::   |||:::
Db       1376 ThrLysTyrGluThrAspAlaIleGlnArgThrGluGluLeuGluGluAlaLys------ 1393

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGAT---GAAGAAAACATA------------ 1152
              |||   :::   |||             ||||||   ||||||::::::
Db       1394 LysLysLeuAlaGln------ArgLeuGlnAspAlaGluGluHisValGluAlaValAsn 1411

Qy       1153 --------------------------------GGAATTTATAAAGAACTAGAAGATTTG 1179
                                                    :::      |||:::|||||||||
Db       1412 AlaLysCysAlaSerLeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeu 1431

Qy       1180 ATA---------------------------------------GAGAAAAATGAAAAT 1197
              :::                                      :::|||      |||
Db       1432 MetIleAspValGluArgThrAsnAlaAlaCysAlaAlaLeuAspLysLysGlnArgAsn 1451

Qy       1198 TTAGATGATTTA-----------------------------------------GATGAA 1215
                 |||   :::                                         |||
Db       1452 PheAspLysIleLeuAlaGluTrpLysHisLysTyrGluGluThrHisAlaGluLeuGlu 1471

Qy       1216 GGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
                 :::|||   |||       ||||||   |||       |||             |||
Db       1472 AlaSerGlnLysGluSerArgSerLeuSerThrGluValPheLysValLysAsnAlaTyr 1491

Qy       1276 GAAAAAACAAAGGAT------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
              |||::::::   |||                |||      :::|||:::|||      |||
Db       1492 GluGluSerLeuAspGlnLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGlu 1511

Qy       1330 ------------------------------------------------CATATTAAAAAA 1341
                                                              ::::::|||
Db       1512 IleSerAspLeuThrGluGlnIleAlaGluGlyGlyLysArgIleHisGluLeuGluLys 1531

Qy       1342 TATAAAAATGATAAGCAGGTTAATAAGGAAAAG--------------------GAAAAA 1380
                 |||       |||||||||   :::||||||                  |||:::
Db       1532 ValLys------LysGlnValGluGlnGluLysSerGluLeuGlnAlaAlaLeuGluGlu 1549

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
                    ||||||   |||            :::   :::|||||||:::|||
Db       1550 AlaGluAlaSerLeuGluHis---------GluGluGlyLysIleLeuArgIleGlnLeu 1566

Qy       1441 GAGTTATCTGAA 1452
              ||||||::::::
```

```
Db       1567 GluLeuAsnGln 1570

RESULT 21
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: UNIPROT:Q07569; UNIPARC:UPI00000802D3; EMBL:L03534; NID:g1850912; PI
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Alignment Scores:
Pred. No.:              8.8e-10       Length:         2139
Score:                  313.50        Matches:        128
Percent Similarity:     46.20%        Conservative:   127
Best Local Similarity:  23.19%        Mismatches:     184
Query Match:            12.61%        Indels:         113
DB:                     2             Gaps:           25

US-09-837-344-45 (1-1482) x T18296 (1-2139)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT----------------------- 36
              :::|||::::::   ||||||||||||:::|||
Db       1014 GluGluGluLysAspAspLeuGluGlnAspArgAlaAspValSerAlaThrLysAspAsp 1033

Qy         37 -------------------------------------------GCTAAAGAAAAGTTGCAA 54
                                                         ||||||:::::::::
Db       1034 IleAlaLysLysLeuAsnLysIleThrIleGluCysGluAspAlaLysAspGluIleAla 1053

Qy         55 GAACAACAAAGCGATTTAGAACAAGATAGA---------------CTTGCTAAAGAA--- 96
              :::   :::   :::||||||   :::                  |||      |||
Db       1054 LysLeuGluGlnGluLeuGluAspGluGluAsnLysAsnLysAspLeuThrAsnGluLeu 1073

Qy         97 ---------------AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACTTGCT 141
                            ||||||   |||   :::      |||      :::      |||
Db       1074 GlnGlnThrGlnLeuLysLeuGlyGluThrGluLysSerLeuAlaAlaGlnValAlaAla 1093

Qy        142 AAAGAAAAGTTGCAAGAACAACAA------------AGCGATCTAGAACAAGAGAGACGT 189
                ::|||      :::::::::                :::||||||   |||:::
Db       1094 ThrLysLysAlaSerAspGluArgAspThrLeuSerGlnAsnLeuGluAsnGluLysLeu 1113

Qy        190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG------------AGA 237
                  :::   |||   :::   ::::::||||||||||::::::             :::
Db       1114 ThrThrLysAsnLeuThrLysThrLysAlaAspLeuGluLysLysIleSerGlyLeuLys 1133

Qy        238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCT 294
              :::      |||   |||::::::   ::::::   :::|||   |||   |||   |||
Db       1134 GlnAspTyrGluAspLeuGluAspAspLysAsnLysIleGluGlyAspLeuArgAsnAla 1153

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 354
              :::   |||::::::|||         ::::::   :::                 :::
Db       1154 GlnArgLysIleLysGluLeuAspAspGluIleThrLysGlyAlaAspValSerGlnTyr 1173

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
              ||||||:::|||:::   :::   |||   :::                |||:::||||||
Db       1174 LeuGlnLysGlnLysGluGluTyrGluSerGln---------IleAlaLysMetGlnGlu 1190

Qy        415 CAACAAAGCGATTTA---------------GAACAAGAGAGACTTGCTAAAGAAAAG--- 456
              ::::::      :::                    :::|||:::      |||||||||
Db       1191 GluLysGluAlaIleGlyAsnAspValLysAsnLysGluLysThrIleLysGluLysGlu 1210

Qy        457 ------TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 510
                    :::|||      |||      |||::::::            ||||||

Db       1211 LeuGluIleGlnSerLeuGlnGluLysLeuAspGluThrGluValGluLysGluAspAla 1230

Qy        511 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 570
              ::::::::::::   ::::::|||:::|||   :::|||         |||||||||:::
Db       1231 GluLysLysLysLysGluIleGluLysGluMetLysAla---------LeuGlnGluGlu 1247

Qy        571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
              :::   ::::::|||      :::   :::   ||||||           ::::::
Db       1248 LysGluAsnValGluSerSerLysAsnSerThrGluLys---------AspLysLysLys 1264

Qy        631 TTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGAT 690
              ||||||            ||||||:::|||   |||:::                  :::
Db       1265 LeuGluAspAsnLeuLysAspThrGlnLysLysLeuAspAspMetThrAlaAspAsnGlu 1284

Qy        691 ATATTA---------GCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTT 741
                 |||         |||:::||||||      :::|||                |||:::
Db       1285 LysLeuLysAlaLysAlaLysAspLeuGluAlaGlnLeuAsn-----------GluVal 1300

Qy        742 CCATCAGAAAATGAACGTGGATAT---------TATATACCACATCAATCTTCTTTACCT 792
                      :::|||:::                    :::   ::::::
Db       1301 GlnAspAsnHisGluLysAlaValAlaAspAlaGluLeuLeuAsnLysLysLysAlaGln 1320

Qy        793 CAGGACAACAGAGGGAATAGTAGAGAT-----------------TCCAAGGAAATGTCT 834
              |||         ||||||                        :::|||      |||
Db       1321 SerAspLysGluLeuAsnSerLeuLysAlaGluLeuGluAlaLeuThrLysAlaLysSer 1340

Qy        835 ATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATA 894
              ::::::|||      |||::::::|||                           :::
Db       1341 ValValGluSerLysAsnLysAspSer-------------------------GluAsn 1351

Qy        895 CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGAT 954
              |||      |||   |||:::   |||   :::         |||:::   ::::::
Db       1352 GluLysAlaAlaLeuSerGluGluIleAspGlnAla------AsnGluLysLeuLysAsn 1369

Qy        955 AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
                    ||||||::::::               |||   ::::::   :::   |||:::
Db       1370 IleGlnAlaAspLeuArgLys---------AlaThrAlaAspLeuGlnGluAlaAsnGlu 1386

Qy        1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGAT 1074
                          :::   |||:::   |||:::   |||      :::   |||
Db       1387 --------------LysLysAlaGluValGluAlaGlnArgAspLysLeuValAlaAsp 1401

Qy        1075 GAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTC 1134
               ::::::      :::   |||:::|||   |||      :::   :::   :::
Db       1402 AsnLysLysMetThrLysThrLeuGluGluIleLysAlaArgAspGluGluAsnThrTyr 1421

Qy        1135 CAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAA---GATTTGATAGAGAAAAAT 1191
               :::   |||              :::   |||      |||   ||||||   |||   |||
Db       1422 LysValGluAsnTyrGluLysValLeuLysArgLysGluAlaAspLeuGluGluAlaAsn 1441

Qy        1192 GAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA 1251
               ||||||||||||               |||||||||                  |||:::
Db       1442 GluAsnLeuAsp--------------IleGluLysLysAspArgMetAsnLysGluLys 1456

Qy        1252 AAAATAAAAAAAGGAAAG---AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
               ::::::|||||   :::   :::   ::::::|||||||||          |||   |||
Db       1457 GlnValLysLysLeuGluGlyGluLeuLysGluThrLysAsp---------LysLeuAsn 1473

Qy        1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                          :::   |||      :::         ||||||   :::
Db       1474 AlaAlaIleAlaGluLysAspSerIle------PheThrAlaLysLysGlnSerAspAla 1491

Qy        1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
               :::   |||:::      |||::::::                           |||
Db       1492 AspLeuGluGluLeuAsnLysThrVal--------------------GluGluHisAsp 1504

Qy        1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
               :::::::|||   :::|||:::      |||||||||
Db       1505 GluValValAlaLysLeuAsnThrGlnIleThrLys 1516

RESULT 22
A29320
myosin heavy chain, fast skeletal muscle, embryonic [similarity] - chicken
```

```
N;Alternate names: myosin heavy chain, EFW1
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C;Accession: A29320; B24124; A02990
R;Molina, M.I.; Kropp, K.E.; Gulick, J.; Robbins, J.
J. Biol. Chem. 262, 6478-6488, 1987
A;Title: The sequence of an embryonic myosin heavy chain gene and isolation of its corre
A;Reference number: A29320; MUID:87194881; PMID:3571266
A;Accession: A29320
A;Molecule type: DNA; mRNA
A;Residues: 1-1940 <MOL>
A;Cross-references: UNIPROT:P02565; UNIPARC:UPI00001775FE
A;Note: the sequence translated from the mRNA differs from that of the DNA in having 379
R;Gulick, J.; Kropp, K.; Robbins, J.
J. Biol. Chem. 260, 14513-14520, 1985
A;Title: The structure of two fast-white myosin heavy chain promoters. A comparative stu
A;Reference number: A92507; MUID:86033956; PMID:2997212
A;Accession: B24124
A;Molecule type: DNA
A;Residues: 1-168 <GUL>
A;Cross-references: UNIPARC:UPI00001775FF
R;Kavinsky, C.J.; Umeda, P.K.; Sinha, A.M.; Elzinga, M.; Tong, S.W.; Zak, R.; Jakovcic,
J. Biol. Chem. 258, 5196-5205, 1983
A;Title: Cloned mRNA sequences for two types of embryonic myosin heavy chains from chick
A;Reference number: A02990; MUID:83161144; PMID:6833296
A;Accession: A02990
A;Molecule type: mRNA
A;Residues: 1502-1546,'A',1548-1912,'G',1914,'T',1916-1940 <KAV>
A;Cross-references: UNIPARC:UPI000017136C; GB:V00430; GB:J00892; NID:g63619; PIDN:CAA237
A;Experimental source: clone 251
C;Comment: The entire sequence of this fragment corresponds to the light meromyosin port
ides (heptads I-IV), that are characteristic of coiled coils.
C;Comment: Association of the alpha-helical regions of two myosin heavy chains to form a
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; methylated amino acid; muscle; muscle contr
F;89-771/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;130/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted

Alignment Scores:
Pred. No.:                9.51e-10      Length:          1940
Score:                    313.00        Matches:         162
Percent Similarity:       41.27%        Conservative:    124
Best Local Similarity:    23.38%        Mismatches:      181
Query Match:              12.59%        Indels:          226
DB:                       2             Gaps:            38

US-09-837-344-45 (1-1482) x A29320 (1-1940)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
              |||   :::   ::::::     |||     |||||:::   ||||||:::::::::
Db        932 GluAspGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 951

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
                 |||:::|||::::::|||                          ::::::   |||
Db        952 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuThrLeuAlaLysValGluLys 971

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
              ||||||                         |||   |||:::   :::   |||:::
Db        972 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaAlaLeuAsp 991

Qy        127 CAA------GAGAGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
              :::      :::|||   |||||||||      |||||||||      ||||||:::
Db        992 GluThrIleAlaLysLeuThrLysGluLysLysAlaLeuGlnGluAlaHisGlnGlnThr 1011

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
                  ||||||:::   |||                        :::||||||   |||
Db       1012 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuThrLysAlaLysThrLys 1031

Qy        202 TTGCAAGAACAACAAAGCGAT------------TTAGAACAAGAGAGA------------ 237
              |||::::::|||      |||            ||||||||||||:::
Db       1032 LeuGluGlnGlnValAspAspLeuGluGlySerLeuGluGlnGluLysLysLeuArgMet 1051

Qy        238 ---------CGTGCTAAAGAAAAGTTG--------------------CAAGAACAACAA 267
                      |||||||||   ||||||                    ||||||
Db       1052 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysMetThrGlnGluSerThr 1071

Qy        268 AGCGATTTAGAACAAGATAGA---CTTGCTAAAGAAAAGTTACAAGAGCAG--------- 315
              |||||||||   |||:::           |||||||||:::::::::
Db       1072 MetAspLeuGluAsnAspLysGlnGlnLeuAspGluLysLeuLysLysLysAspPheGlu 1091

Qy        316 ------------CAAAGCGATTTAGAACAAGAG-----------------------AGA 339
                         ||||||   :::|||   |||                       :::
Db       1092 IleSerGlnIleGlnSerLysIleGluAspGluGlnAlaLeuGlyMetGlnLeuGlnLys 1111

Qy        340 CGTGCTAAAGAA---------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 390
              :::   ||||||         :::::::::|||   :::   ::::::|||   ||||||
Db       1112 LysIleLysGluLeuGlnAlaArgIleGluGluLeuGluGluGluIleGluAlaGluArg 1131

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG---AGACTTGCT 447
                 ::::::   |||   ::::::   ::::::||||||   :::|||
Db       1132 ThrSerArgAlaLysAlaGluLysHisArgAlaAspLeuSerArgGluLeuGluGluIle 1151

Qy        448 AAAGAAAAGTTGCAAGAA-----------------CAACAAAGCGATTTAGAACAAGAG 489
              |||:::|||:::|||                    |||   |||:::   ::::::
Db       1152 SerGluArgLeuGluGluAlaGlyGlyAlaThrAlaAlaGlnIleAspMetAsnLysLys 1171

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------- 540
              |||   |||      :::   |||:::   :::   |||||||||:::
Db       1172 ArgGluAla------GluPheGlnLysMetArgArgAspLeuGluGluAlaThrLeuGln 1189

Qy        540 ------------------------------------------------------------ 540

Db       1190 HisGluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerThrAlaAspValGly 1209

Qy        541 -----------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
                              :::|||   |||:::||||||:::::::::::|||:::|||
Db       1210 GluGlnIleAspAsnLeuGlnArgValLysGlnLysLeuGluLysGluLysSerGluLeu 1229

Qy        583 GAACAAGAG--------------------------------AGACGTGCTAAAGAAAAG 609
              :::   |||                                :::||||||
Db       1230 LysMetGluIleAspAspLeuAlaSerAsnMetGluSerValSerLysAlaLysAlaAsn 1249

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAA---AAAAATTTA 666
              |||::::::      |||   ||||||   :::   ::::::   |||   |||
Db       1250 LeuGluLysMetCysArgSerLeuGluAspGlnLeuSerGluIleLysThrLysGluGlu 1269

Qy        667 GAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATA 726
              |||::::::::::         ||||||   |||:::            ||||||:::
Db       1270 GluGlnGlnArgThrIleAsnAspIleSerAlaGlnLys------AlaArgLeuGlnThr 1287

Qy        727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786
                 :::   |||         :::   |||                          :::
Db       1288 GluSerGlyGluTyrSerArgGlnValGluGlu---------------LysAspAlaLeu 1302

Qy        787 TTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAA 846
              :::   |||   :::||||||           ::::::::::::      :::::::::
Db       1303 IleSerGlnLeuSerArgGlyLysGlnAlaPheThrGlnGlnIleGluGluLeuLysArg 1322

Qy        847 ACAAATAGAGAATCTATTACAACAAAT--------------------GTTGAAGGACGA 885
                 |||   |||                                    |||
Db       1323 HisLeuGluGluGluIleLysAlaLysLysCysProAlaHisAlaLeuGlnSerAlaArg 1342

Qy        886 AGGGATATACAT-----------------------------------AAAGGACATCTT 909
                 |||                                        ||||||   |||
Db       1343 HisAspCysAspLeuLeuArgGluGlnTyrGluGluGluGlnGluAlaLysGlyGluLeu 1362

Qy        910 GAAGAA-------------------------------------------AAGAAA 921
              :::                                              :::
Db       1363 GlnArgAlaLeuSerLysAlaAsnSerGluValAlaGlnTrpArgThrLysTyrGluThr 1382

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGAT--------------------AAATCT 960
              |||         :::   |||:::   |||:::                       :::
Db       1383 AspAlaIleGlnArgThrGluGluLeuGluGluAlaLysLysLysLeuAlaGlnArgLeu 1402

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
```

```
Db       1403 GlnAspAlaGluGluHis---ValGluAlaValAsnSerLysCysAlaSer--------- 1418

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
Db       1419 LeuGluLysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetIleAspValGlu 1438

Qy       1081 GAAGAT----------------------------------------------GATGAAGACTTA 1098
Db       1439 ArgSerAsnAlaAlaCysAlaAlaLeuAspLysLysGlnLysAsnPheAspLysIleLeu 1458

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA--------------- 1143
Db       1459 SerGluTrpLysGln------LysTyrGluGluThrGlnAlaGluLeuGluAlaSerGln 1476

Qy       1144 --------------------------------GAAAACATAGGAATTTATAAAGAA--- 1167
Db       1477 LysGluSerArgSerLeuSerThrGluLeuPheLysMetLysAsnAlaTyrGluGluSer 1496

Qy       1168 ---------CTAGAAGATTTGATAGAGAAAAATGAAAATTTA------------GATGAT 1206
Db       1497 LeuAspHisLeuGluThrLeuLysArgGluAsnLysAsnLeuGlnGlnGluIleSerAsp 1516

Qy       1207 TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA---GAAAAAATAAAAAAA 1263
Db       1517 LeuThrGluGlnIleAlaGluGlyGlyLysAlaIleHisGluLeuGluLysValLysLys 1536

Qy       1264 GGAAAGAAATATGAAAAAACAAAG---GATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
Db       1537 GlnIleGluGlnGluLysSerGluLeuGlnThrAlaLeuGluGluAlaGluAlaSerLeu 1556

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1557 GluHisGluGluGlyLysIleLeuArgValGlnLeuGluLeuAsnGlnValLysSerAsp 1576

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT--------------- 1425
Db       1577 IleAspArgLys---------IleAlaGluLysAspGluGluIleAspGlnLeuLysArg 1593

Qy       1426 ------TTACAGATCGTGGATGAGTTATCTGAAGATATA 1458
Db       1594 AsnHisLeuArgValValAspSerMetGlnSerThrLeu 1606
```

RESULT 23
A59234
slow myosin heavy chain 3 - quail
C;Species: Coturnix coturnix
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C;Accession: A59234
R;Nikovits Jr., W.; Wang, G.F.; Feldman, J.L.; Miller, J.B.; Wade, R.; Nelson, L.; Stock
J. Biol. Chem. 271, 17047-17056, 1996
A;Title: Isolation and characterization of an avian slow myosin heavy chain gene expres
A;Reference number: A59234; MUID:96291845; PMID:8663323
A;Accession: A59234
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1931 <NIK>
A;Cross-references: UNIPARC:UPI00000FB7EB; GB:U53862; NID:g1289513; PIDN:AAC59912.1; PID
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;81-761/Domain: myosin motor domain homology <MMO>

Alignment Scores:

Pred. No.:	1.07e-09	Length:	1931
Score:	312.00	Matches:	153
Percent Similarity:	41.79%	Conservative:	127
Best Local Similarity:	22.84%	Mismatches:	192
Query Match:	12.55%	Indels:	198
DB:	2	Gaps:	33

US-09-837-344-45 (1-1482) x A59234 (1-1931)

```
Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
Db        922 GluGluGluGluGluMetAsnAlaGluLeuThrAlaLysLysArgLysLeuGluAspGlu 941

Qy         61 CAAAGCGATTTAGAACAAGAT--------------------------AGACTTGCTAAA 93
Db        942 CysSerGluLeuLysLysAspIleAspAspLeuGluLeuSerLeuAlaLysValGluLys 961

Qy         94 GAAAAG--------------------------TTACAAGAGCAGCAAAGCGATTTAGAA 126
Db        962 GluLysHisAlaThrGluAsnLysValLysAsnLeuThrGluGluMetAlaGlyLeuAsp 981

Qy        127 CAAGAG------AGACTTGCTAAAGAAAAG------TTGCAAGAA------CAACAAAGC 168
Db        982 GluAsnIleThrLysLeuThrLysGluLysLysThrLeuGlnGluSerHisGlnGlnAla 1001

Qy        169 ------GATCTAGAACAAGAGAGA--------------------CGTGCTAAAGAAAAG 201
Db       1002 LeuAspAspLeuGlnAlaGluGluAspLysValAsnThrLeuAlaLysAlaLysValLys 1021

Qy        202 TTGCAAGAACAACAAAGCGATTTAGAA------------CAAGAGAGACGTGCTAAAGAA 249
Db       1022 LeuGluGlnGlnAlaAspAspLeuGluSerSerLeuGlnGlnGluLysLysIleArgMet 1041

Qy        250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTA 306
Db       1042 AspLeuGluArgAlaLysArgLysLeuGluGlyAspLeuLysLeuAlaGlnGluSerVal 1061

Qy        307 CAAGAGCAGCAAAGCGATTTAGAACAA---GAGAGACGTGCTAAAGAAAAG--------- 354
Db       1062 MetAspLeuGluAsnAspLysGlnGlnLeuGluGluArgLeuLysLysLysAspPheGlu 1081

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
Db       1082 LeuAsnThrLeuAsnAlaArgIleGluAspGluGlnAlaIleAlaAlaGlnLeuGlnLys 1101

Qy        415 CAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
Db       1102 LysLeuLysGluLeuGln-----------------AlaArgIleGluGluLeuGluGlu 1115

Qy        475 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 534
Db       1116 GluLeuGluAlaGluArgThrGlyArgAlaLysValGluLysLeuArgSerGluLeuLeu 1135

Qy        535 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAA------------CAACAAAGCGAT 579
Db       1136 GlnGluLeuGluGluThrSerGluArgLeuGluGluAlaGlyGlyAlaThrSerValGln 1155

Qy        580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
Db       1156 LeuGluLeuAsnLysLysArgGluAlaGluPheGlnLysLeuArgArgAspLeuGluGlu 1175

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGA---AAAAAGGAACATGGAGATATATTA 696
Db       1176 AlaThrLeuGlnHisGluAlaThrAlaAlaAlaLeuArgLysLysHisAlaAspSerVal 1195

Qy        697 GCAGAGGATTTATATGGTCGTTTAGAA---ATACCAGCTATAGAACTTCCATCAGAAAAT 753
Db       1196 AlaGlu---LeuSerGluGlnLeuAspAsnMetGlnArgValLysGlnLysLeuGluLys 1214

Qy        754 GAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGT 813
Db       1215 GluLysSer---------------GluLeuLysLeuGluLeuAspAspValSerSer--- 1228

Qy        814 AGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAAT 873
Db       1229 ------AsnMetGluGlnLeuIleLysAlaLysAlaAsnLeuGluLysMetCysArgSer 1246

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAA------------ 921
Db       1247 ThrGluAspGlnMetAsnGluHisArgAsnLysLeuGluGluSerGlnArgThrValThr 1266

Qy        922 --------------------------------GATGGTTCAATAAAACCAGAACAAAAA 948
Db       1267 AspLeuSerThrGlnArgAlaLysLeuGlnThrGluAsnSerGluLeuSerArgGlnLeu 1286
```

```
Qy        949 GAAGATAAATCTGCTGACATA-----------CAAAATCATACATTAGAGACAGTAAAT 996
              |||:::|||   |||   |||           :::                 |||
Db       1287 GluGluLysGluAlaPheIleAsnGlnLeuMetArgGlyLysLeuThrTyrThrGlnGln 1306

Qy        997 ATTTCTGATGTTAAT------------------------------------------- 1011
              :::   |||:::
Db       1307 LeuGluAspLeuLysArgGlnLeuGluGluGluAlaLysAlaLysAsnAlaLeuAlaHis 1326

Qy       1012 --------------------------GATTTTCAAATAAGTAAGTATGAGGATGAAATA 1044
                                         |||           :::|||||:::|||:::
Db       1327 AlaLeuGlnSerAlaGlnHisAspCysAspLeuLeuArgGluGlnTyrGluGluGluMet 1346

Qy       1045 AGTGCT---------------------------------------------------- 1050
                 |||
Db       1347 GluAlaLysAlaGluLeuGlnArgAlaLeuSerLysAlaAsnSerGluValAlaGlnTrp 1366

Qy       1051 ------GAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGAC--------- 1095
                  :::|||:::         |||         |||:::   |||:::
Db       1367 ArgThrLysTyrGluThrAspAlaIleGlnArgThrGluGluLeuGluGluAlaLysLys 1386

Qy       1096 --------------TTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAA--- 1137
                           |||   |||   :::     |||:::     |||   :::
Db       1387 LysLeuAlaGlnArgLeuGlnGluAlaGluGluAlaValGluAlaValAsnAlaLysCys 1406

Qy       1138 ---------GATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATA------ 1182
                      ::::::          :::      |||:::||||||||:::
Db       1407 SerSerLeuGluLysThrLysHisArgLeuGlnAsnGluIleGluAspLeuMetAlaAsp 1426

Qy       1183 ---------------------------------GAGAAAAATGAAAATTTAGATGAT 1206
                                                :::|||      |||   |||
Db       1427 ValGluArgSerAsnAlaAlaAlaAlaAlaLeuAspLysLysGlnArgAsnPheAspLys 1446

Qy       1207 TTA-------------------------------------------GATGAAGGAATAGAA 1224
              :::                                                |||     :::
Db       1447 IleLeuSerGluTrpLysGlnLysPheGluGluSerGlnThrGluLeuGluAlaSerGln 1466

Qy       1225 AAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACA 1284
              |||   :::      ||||||   |||      |||             ||||||::::::
Db       1467 LysGluAlaArgSerLeuSerThrGluLeuPheLysLeuLysAsnAlaTyrGluGluSer 1486

Qy       1285 AAGGATAAT------AATTTTAAACCAAATGATAAAAGTTTGTATGATGAG---------- 1329
                 ::::::         ||||||      :::|||:::|||   :::|||
Db       1487 LeuGluHisLeuGluThrPheLysArgGluAsnLysAsnLeuGlnGluGluIleSerAsp 1506

Qy       1330 ---------------------------------------CATATTAAAAAATATAAAAAT 1350
                                                       ::::::|||   :::
Db       1507 LeuThrGluGlnLeuGlyAlaSerGlnLysSerIleHisGluLeuGluLysValArg--- 1525

Qy       1351 GATAAGCAGGTTAATAAGGAAAAG--------------------GAAAAATTCATAAAA 1389
                 ||||||::::::   ||||||              |||:::
Db       1526 ---LysGlnLeuAspAlaGluLysLeuGluLeuGlnAlaAlaLeuGluGluAlaGluAla 1544

Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
              ||||||   |||            :::   :::||||||:::         |||   :::
Db       1545 SerLeuGluHis---------GluGluGlyLysIleLeuArgAlaGlnLeuGluPheAsn 1561

Qy       1450 GAAGATATAACTAAATATTTTATGAAACTA 1479
              :::          :::|||     ||||||
Db       1562 GlnValLysAlaGluTyrGluArgLysLeu 1571

RESULT 24
B43402
myosin heavy chain-B, neuronal - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Dec-2004
C;Accession: B43402; A43402
R;Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A;Reference number: A43402; MUID:92388144; PMID:1355479
A;Accession: B43402
A;Molecule type: mRNA
A;Residues: 1-2007 <TAK>
A;Cross-references: UNIPROT:Q02015; UNIPARC:UPI00000FC6BE; GB:M93676; NID:g212448; PIDN
A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide s
A;Accession: A43402
A;Molecule type: mRNA
A;Residues: 1-211;222-631;653-2007 <TA2>
A;Cross-references: UNIPARC:UPI0000173E32; UNIPARC:UPI0000173E33; UNIPARC:UPI0000173E34
A;Note: sequence extracted from NCBI backbone (NCBIN:112864)
C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle my
C;Superfamily: myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylate
F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F;1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS
F;88-802/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;212-221/Region: alternatively spliced segment 1 #status experimental
F;559-593/Region: actin binding #status predicted
F;632-652/Region: alternatively spliced segment 2 #status experimental
F;692-714/Region: actin binding #status predicted
F;875-2007/Domain: coiled coil #status predicted <COI>
F;875-1315/Region: S2
F;1316-2007/Region: light meromyosin
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;732,742/Active site: Cys #status predicted
F;1954/Binding site: phosphate (Thr) (covalent) #status predicted
F;1987/Binding site: phosphate (Ser) (covalent) #status predicted

Alignment Scores:
Pred. No.:              1.07e-09        Length:         2007
Score:                  312.00          Matches:        136
Percent Similarity:     44.84%          Conservative:   138
Best Local Similarity:  22.26%          Mismatches:     163
Query Match:            12.55%          Indels:         174
DB:                     1               Gaps:           28

US-09-837-344-45 (1-1482) x B43402 (1-2007)

Qy         10 CAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 69
              :::::::::||||||:::       |||       ::::::|||   |||:::::::::
Db        903 GluAlaGluLeuGluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIle 922

Qy         70 TTAGAACAAGAT------------------------------------------------ 81
              |||   :::
Db        923 LeuAlaGluGlnLeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAla 942

Qy         82 AGACTTGCT------AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTA------------ 123
              |||||||||      |||::::::|||:::|||         ||||||
Db        943 ArgLeuAlaAlaLysLysGlnGluLeuGluGluIleLeuHisAspLeuGluSerArgVal 962

Qy        124 ------GAACAAGAGAGA---------CTTGCTAAAGAAAAG------------------ 150
                    |||:::||||||         |||      ||||||
Db        963 GluGluGluGluGluArgAsnGlnIleLeuGlnAsnGluLysLysLysMetGlnGlyHis 982

Qy        151 TTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 210
              :::|||:::   :::      |||::::::|||      |||::::::|||||||||
Db        983 IleGlnAspLeuGluGluGlnLeuAspGluGluGluGlyAlaArgGlnLysLeuGlnLeu 1002

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG------------TTGCAA 258
              ::::::         |||   :::   :::      |||:::             |||:::
Db       1003 GluLysValThrAlaGluAlaLysIleLysLysMetGluGluGluIleLeuLeuLeuGlu 1022

Qy        259 GAACAACAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
              :::|||   |||      |||::::::   :::|||   ::::::::::::   |||
Db       1023 AspGlnAsnSerLysPheLeuLysGluLysLysLeuMetGluAspArgIleAlaGluCys 1042

Qy        316 CAAAGCGATTTA---GAACAAGAGAGACGTGCTAAA---------------------- 348
                 |||   |||   |||:::|||   :::||||||
Db       1043 ThrSerGlnLeuAlaGluGluGluGluLysAlaLysAsnLeuAlaLysLeuLysAsnLys 1062

Qy        349 --------------------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 381
```

```
                                        |||:::|||:::::::::::::         :::
Db       1063 GlnGluMetMetIleThrAspLeuGluGluArgLeuLysLysGluGluLysThrArgGln 1082

Qy        382 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGA 441
              :::      :::||||||   ||||||      :::   :::||||||   |||::::::
Db       1083 GluLeuGluLysAlaLysArgLysLeuAspGlyGluThrThrAspLeu---GlnAspGln 1101

Qy        442 CTTGCTAAA--------------GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 486
              :::|||:::               |||:::|||:::   |||   :::       |||:::
Db       1102 IleAlaGluLeuGlnAlaGlnIleGluGluLeuLysIleGlnLeuAlaLysLysGluGlu 1121

Qy        487 GAG--------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 531
              |||                |||       |||:::      ::::::   :::   |||
Db       1122 GluLeuGlnAlaAlaLeuAlaArgGlyAspGluGluAlaValGlnLysAsnAsnAlaLeu 1141

Qy        532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 591
              :::      |||      :::   ::::::   |||   |||   |||||||||   |||
Db       1142 LysValIleArgGluLeuGlnAlaGlnIleAlaGluLeuGlnGluAspLeuGluSerGlu 1161

Qy        592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTA------------------ 633
              :::   ::::::   |||   ::::::|||:::|||||||||
Db       1162 LysAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuSerGluGluLeuGluAla 1181

Qy        633 ------------------------------------------------------------ 633

Db       1182 LeuLysThrGluLeuGluAspThrLeuAspThrThrAlaAlaGlnGlnGluLeuArgThr 1201

Qy        634 ------GAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687
                    ||||||      |||:::   ||||||   :::|||   :::   ::::::
Db       1202 LysArgGluGlnGluValAlaGluLeuLysLysAlaIleGluGluGluThrLysAsnHis 1221

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
              :::         :::::::::      |||                  ||||||
Db       1222 GluAlaGlnIleGlnGluIleArgGlnArgHisAlaThrAlaLeuGluGluLeuSerGlu 1241

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
              :::   |||:::                 :::::::::|||   ::::::   :::|||
Db       1242 GlnLeuGluGlnAlaLysArgPhe------LysAlaAsnLeuGluLysAsnLysGlnGly 1259

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCT------------ATAATAGAAAAAACAAATAGA 855
                    |||:::||||||:::::           ::::::::::
Db       1260 LeuGluSerAspAsnLysGluLeuAlaCysGluValLysValLeuGlnGlnValLysAla 1279

Qy        856 GAATCT---------------------------------------ATTACAACAAATGTT 876
              ||||||                                      :::|||      |||
Db       1280 GluSerGluHisLysArgLysLysLeuAspAlaGlnValGlnGluLeuThrAlaLysVal 1299

Qy        877 ---GAAGGACGAAGG---------------------------------------------- 888
                 ||||||   |||
Db       1300 ThrGluGlyGluArgLeuArgValGluLeuAlaGluLysAlaAsnLysLeuGlnAsnGlu 1319

Qy        889 --------------------------GATATACATAAAGGACATCTTGAAGAAAAGAAA 921
                                       :::      |||      ::::::      |||
Db       1320 LeuAspAsnValSerSerLeuLeuGluGluAlaGluLysLysGlyIleLysPheAlaLys 1339

Qy        922 GATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
              |||   :::              |||   :::|||            :::|||      |||
Db       1340 AspAlaAlaSerLeuGluSerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThr 1359

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
                 :::   :::|||:::|||               :::|||   :::   |||:::|||
Db       1360 ArgGlnLysLeuAsnLeuSerSer-----------ArgIleArgGlnLeuGluGluGlu 1375

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
                          :::::::|||   :::::::::|||::::::|||:::
Db       1376 --------------LysAsnAsnLeuGlnGluGlnGlnGluGluGluGluAlaArg 1390

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
              :::                        |||      :::::::::   :::   :::
Db       1391 Lys----------------------AsnLeu---GluLysGlnMetLeuAlaLeuGln 1401

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT------GAA 1215
                 :::|||   :::      :::|||   :::      |||||||||         |||
```

```
Db       1402 AlaGlnLeuAlaGluAlaLysLysLysValAsp------AspAspLeuGlyThrIleGlu 1419

Qy       1216 GGAATAGAA--------------------AAATCATCAGAAGAATTATCT--------- 1245
              |||:::|||                    |||      |||   ||||||
Db       1420 GlyLeuGluGluAsnLysLysLysLeuLeuLysAspMetGluSerLeuSerGlnArgLeu 1439

Qy       1246 GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
              |||||||||               |||   ||||||||||||:::         :::
Db       1440 GluGluLysAlaMetAlaTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnGlu 1459

Qy       1306 AATGATAAAAGTTTGTATGAT---GAGCATATTAAAAAATATAAAAATGATAAGCAGGTT 1362
                 |||      :::   |||   :::|||   ::::::      ::::::      :::
Db       1460 LeuAspAspLeuMetValAspLeuAspHisGlnArgGlnIleValSerAsn------Leu 1477

Qy       1363 AATAAGGAAAAGGAAAAATTCATAAAATCATTG 1395
                 |||:::::::::|||||   :::   |||
Db       1478 GluLysLysGlnLysLysPheAspGlnMetLeu 1488

RESULT 25
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Dictyostelium discoideum
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C;Accession: A26655; A24728; S00250
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
A;Reference number: A26655; MUID:87092266; PMID:3540939
A;Accession: A26655
A;Molecule type: DNA
A;Residues: 1-2116 <WAR>
A;Cross-references: UNIPROT:P08799; UNIPARC:UPI000012FBD3; GB:M14628; GB:M11938; NID:g16
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A;Reference number: A24728; MUID:86016788; PMID:3901008
A;Accession: A24728
A;Molecule type: mRNA
A;Residues: 2035-2116 <DEL>
A;Cross-references: UNIPARC:UPI0000177634
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
A;Reference number: S00250; MUID:88112226; PMID:2828113
A;Accession: S00250
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1734-1893 <WAG>
A;Cross-references: UNIPARC:UPI0000177635
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos
F;1-818/Domain: globular head <HED>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

Alignment Scores:
Pred. No.:              1.06e-09        Length:         2116
Score:                  312.00          Matches:        130
Percent Similarity:     44.59%          Conservative:   142
Best Local Similarity:  21.31%          Mismatches:     180
Query Match:            12.55%          Indels:         158
DB:                     2               Gaps:           26

US-09-837-344-45 (1-1482) x A26655 (1-2116)

Qy           4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAAGAACAA 60
               :::   :::      ||||||   ::::::|||      |||   ::::::::::::
Db         884 AspSerLysAspAlaLeuGluAlaGlnLysArgGluLeuGluIleArgValGluAspMet 903

Qy          61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
               :::|||:::|||::::::   :::||||||   |||   ||||||   |||:::
```

```
Db        904 GluSerGluLeuAspGluLysLysLeuAlaLeuGluAsnLeuGlnAsnGlnLysArgSer 923

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              :::|||                  ||||||:::::::::   :::   :::|||::::::
Db        924 ValGlu-----------------GluLysValArgAspLeuGluGluGluLeuGlnGlu 937

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              |||::::::   :::       |||::::::   :::          |||:::|||
Db        938 GluGlnLysLeuArgAsnThrLeuGluLysLeuLysLysLysTyrGluGluGluLeuGlu 957

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------------TTAGAA------CAA 282
                  :::::::::   :::   |||||||||             ||||||      :::
Db        958 GluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGluLysIleLys 977

Qy        283 GATAGACTTGCTAAA------GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 336
              |||   |||   |||      |||:::|||   |||      |||:::      ::::::
Db        978 AspGluLeuGlnLysGluValGluGluLeuThrGluSerPheSerGluGluSerLysAsp 997

Qy        337 AGACGTGCTAAAGAAAAG------------TTGCAAGAACAACAAAGCGAT--------- 375
              :::       ||||||             ||||||   :::      |||
Db        998 LysGlyValLeuGluLysThrArgValArgLeuGlnSerGluLeuAspAspLeuThrVal 1017

Qy        376 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 432
                 |||:::   |||   :::   |||   :::|||      |||:::      ||||||
Db       1018 ArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGlnLysLysLysLeuGlu 1037

Qy        433 CAAGAG----------------------------------------AGACTTGCTAAA 450
              :::|||                                        :::||||||:::
Db       1038 GluGluLeuLysGlnValGlnGluAlaLeuAlaAlaGluThrAlaAlaLysLeuAlaGln 1057

Qy        451 GAA------------AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--------- 489
              |||            |||||||||   :::   ::::::|||   ::::::
Db       1058 GluAlaAlaAsnLysLysLeuGlnGlyGluTyrThrGluLeuAsnGluLysPheAsnSer 1077

Qy        490 ------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA------ 519
                                     ::::::|||:::   |||:::   |||
Db       1078 GluValThrAlaArgSerAsnValGluLysSerLysLysThrLeuGluSerGlnLeuVal 1097

Qy        520 ------CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA--------- 564
                  :::::::|||:::::::|||::::::   ::::::   |||:::
Db       1098 AlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAlaLeuGluLysLysLys 1117

Qy        564 ------------------------------------------------------------ 564

Db       1118 LysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGluSerThrGlyGlyGlu 1137

Qy        565 --------------------------GAACAACAAAGCGATTTAGAACAAGAGAGACGT 597
                                       :::|||:::||||||:::|||      |||
Db       1138 LysLysSerLeuTyrAspLeuLysValLysGlnGluSerAspMetGluAlaLeuArgAsn 1157

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGAT------ 651
                  :::||||||                |||||||:::   |||:::
Db       1158 GlnIleSerGluLeuGlnSerThrIleAlaLysLeuGluLysIleLysSerThrLeuGlu 1177

Qy        652 ---------------------------------------------ACGAAAAAAAATTTA 666
                                                          |||   |||:::
Db       1178 GlyGluValAlaArgLeuGlnGlyGluLeuGluAlaGluGlnLeuAlaLysSerAsnVal 1197

Qy        667 GAAAGAAAAAAGGAACATGGAGATATA----------------------TTAGCAGAG 702
              |||::::::|||:::      ::::::                     |||||||||
Db       1198 GluLysGlnLysLysLysValGluLeuAspLeuGluAspLysSerAlaGlnLeuAlaGlu 1217

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
              :::         :::      :::      ::::::         |||   |||
Db       1218 GluThrAlaAlaLysGlnAlaLeuAspLysLeuLysLysLysLeuGluGlnGluLeuSer 1237

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAAT---AGTAGAGAT 819
                     |||:::   |||   :::   |||      |||   :::   |||
Db       1238 GluVal---------GlnThrGlnLeuSerGluAlaAsnAsnLysAsnValAsnSerAsp 1254

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||          :::          |||          :::          |||
Db       1255 SerThrAsnLysHisLeuGluThrSerPheAsnAsnLeuLysLeuGluLeuGluAlaGlu 1274

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT------------ 927
                  :::              |||      ||||||:::|||:::   |||
Db       1275 GlnLysAla---------LysGlnAlaLeuGluLysLysArgLeuGlyLeuGluSerGlu 1291

Qy        928 ------------------TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
                                ::::::   |||:::|||:::   :::|||   :::
Db       1292 LeuLysHisValAsnGluGlnLeuGluGluGluLysLysGlnLysGluSerAsnGluLys 1311

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
              :::          ||||||          :::|||::::::   |||          :::
Db       1312 ArgLysValAspLeuGluLys---GluValSerGluLeuLysAsp-----------Gln 1326

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT--- 1086
              |||:::|||:::::::::::::                  :::   |||   ::::::
Db       1327 IleGluGluGluValAlaSerLys---------LysAlaValThrGluAlaLysAsnLys 1343

Qy       1087 ---GATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA 1143
                 :::   :::|||||||||   |||          ||||||   :::
Db       1344 LysGluSerGluLeuAspGluIleLysArg------GlnTyrAlaAspValValSerSer 1361

Qy       1144 GAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT 1203
                 :::      :::   ::::::|||:::   |||      |||||||||   |||
Db       1362 ArgAspLysSerVal---GluGlnLeuLysThrLeuGlnAlaLysAsnGluGluLeu--- 1379

Qy       1204 GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAA 1263
                                 :::::::::   |||   :::|||   ::::::   :::
Db       1380 --------------------ArgAsnThrAlaGluGluAlaGluGlyGlnLeuAspArg 1392

Qy       1264 GGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTAT 1323
                 ::::::   :::|||      :::   :::   :::          |||:::|||
Db       1393 AlaGluArgSerLysLysLysAlaGluPheAspLeuGluGluAlaValLysAsnLeuGlu 1412

Qy       1324 GATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
              :::|||      ||||||   |||   :::|||   :::   |||   :::      :::
Db       1413 GluGluThrAlaLysLysValLysAlaGluLysAlaMetLysLysAlaGluThrAspTyr 1432

Qy       1384 ATAAAATCATTGTTTCATATATTTGACGGAGACAAT---------------------GAA 1422
                    :::          :::   |||        |||                  :::
Db       1433 ArgSerThrLysSerGluLeuAspAspAlaLysAsnValSerSerGluGlnTyrValGln 1452

Qy       1423 ATTTTACAGATCGTGGATGAGTTATCTGAA 1452
              |||   ::::::   :::||||||||||||
Db       1453 IleLysArgLeuAsnGluGluLeuSerGlu 1462
```

RESULT 26
A54514
glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: GARP
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004
C;Accession: A54514
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Silva, A.; Anders, R.F.; Kemp, D.J.
Mol. Biochem. Parasitol. 31, 199-202, 1988
A;Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich pro
A;Reference number: A54514; MUID:89040048; PMID:2903445
A;Accession: A54514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-678 <TRI>
A;Cross-references: UNIPROT:P13816; UNIPARC:UPI000012B0E0; GB:J03998; NID:g160298; PIDN:
C;Genetics:
A;Introns: 25/3
C;Superfamily: histone H1
C;Keywords: tandem repeat

Alignment Scores:

Pred. No.:	1.53e-09	Length:	678
Score:	310.50	Matches:	113
Percent Similarity:	47.35%	Conservative:	119
Best Local Similarity:	23.06%	Mismatches:	169
Query Match:	12.48%	Indels:	89
DB:	2	Gaps:	16

US-09-837-344-45 (1-1482) x A54514 (1-678)

```
Qy         16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA---CAACAAAGCGATTTA 72
Db        237 AspAsnAspLysGluAsnIleSerGluAspLysLysGluAspHisGlnGlnGluGluMet 256

Qy         73 GAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAG 132
Db        257 LeuLysThrLeuAspLysLysGluArgLysGlnLysGluLysGluMetLysGluGlnGlu 276

Qy        133 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCT 192
Db        277 LysIleGluLysLysLysLysLysGlnGlu-----------GluLysGluLysLysLys 292

Qy        193 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
Db        293 GlnGluLysGluArgLysLysGlnGlu------LysLysGluArgLysGlnLysGluLys 310

Qy        253 TTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 312
Db        311 GluMetLysLysGlnLysLysIleGluLysGluArgLysLysLysGlu------------ 326

Qy        313 CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 372
Db        327 ---------------GluLysGluLysLysLysLysLysHisAspLysGluAsnGluGlu 341

Qy        373 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAG----------------------TTG 408
Db        342 ThrMetGlnGlnProAspGlnThrSerGluGluThrAsnAsnGluIleMetValProLeu 361

Qy        409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
Db        362 ProSerProLeuThrAspValThrThrProGluGluHisLysGluGlyGluHisLysGlu 381

Qy        469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
Db        382 GluGluHisLysGluGlyGluHisLysGluGlyGluHisLysGluGluGluHisLysGlu 401

Qy        529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db        402 ---GluGluHisLysLysGluGluHisLysSerLysGluHisLysSerLysGlyLysLys 420

Qy        589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db        421 AspLysGlyLysLysAspLysGlyLysHisLysLysAlaLysLysGluLysValLysLys 440

Qy        649 GATACGAAAAAA----------------------------------------AATTTA 666
Db        441 HisValValLysAsnValIleGluAspGluAspLysAspGlyValGluIleIleAsnLeu 460

Qy        667 GAAAGAAAAAAG------GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
Db        461 GluAspLysGluAlaCysGluGluGlnHisIleThrValGluSer------ArgProLeu 478

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
Db        479 SerGlnProGlnCysLysLeuIleAspGlu--------------------ProGluGln 491

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
Db        492 LeuThrLeuMetAspLysSerLys---------ValGluGluLysAsnLeuSerIleGln 508

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
Db        509 GluGlnLeuIleGlyThrIleGlyArgValAsnValValProArgArgAspAsnHisLys 528

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATA--------------------AAACCA 939
Db        529 LysLysMetAlaLysIleGluGluAlaGluLeuGlnLysGlnLysHisValAspLysGlu 548

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
```

```
Db        549 GluAspLysLysGluGluSerLysGluValGlnGluGluSerLysGluVal--------- 565

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
Db        566 -----------------------------GlnGluAspGluGluGluValGluGluAsp 575

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
Db        576 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu--------------- 590

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
Db        591 --------------GluGluGluGlu--------------GluGluGluGluGluGlu 600

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
Db        601 GluGluGluGluGluAspGluAspGluGluAspGluAspAspAlaGluGluAspGluAsp 620

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
Db        621 AspAlaGluGluAspGluAspAspAlaGluGluAspAspAspGluGluAspAspAspGlu 640

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
Db        641 GluAspAspAspGluAsp---GluAspGluAspGluGluAspGluGluGluGluGluGlu 659

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
Db        660 GluGluGluGluSerGluLysLysIleLys 669
```

RESULT 27
A33977
myosin heavy chain, nonmuscle - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A33977; S06116; A43422
R;Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A;Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A;Reference number: A33977; MUID:90046668; PMID:2813355
A;Accession: A33977
A;Molecule type: mRNA
A;Residues: 1-1959 <SHO>
A;Cross-references: UNIPROT:P14105; UNIPARC:UPI000012FB7F; GB:M26510; NID:g212382; PIDN:
R;Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A;Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A;Reference number: S06116; MUID:90032648; PMID:2806244
A;Accession: S06116
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 716-1008 <KAT>
A;Cross-references: UNIPARC:UPI0000173E31; GB:X17589
A;Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114
R;Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A;Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate
A;Reference number: A43422; MUID:92381096; PMID:1512291
A;Accession: A43422
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1900-1959 <HOD>
A;Cross-references: UNIPARC:UPI000008A46C
A;Experimental source: brush border
A;Note: sequence extracted from NCBI backbone (NCBIP:111947)
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted

```
F;837-1936/Domain: coiled coil #status predicted <COI>
F;837-1277/Region: S2
F;1278-1959/Region: light meromyosin
F;1937-1959/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694,704/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:                1.28e-09        Length:          1959
Score:                    310.50          Matches:          141
Percent Similarity:       42.98%          Conservative:     110
Best Local Similarity:    24.14%          Mismatches:       164
Query Match:              12.48%          Indels:           169
DB:                       1               Gaps:              27

US-09-837-344-45 (1-1482) x A33977 (1-1959)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||   :::      ||||||:::|||      |||::::::|||||||||   :::
Db        946 GlnGluLeuGluGluGlnLeuGluGluGluGluSerAlaArgGlnLysLeuGlnLeuGlu 965

Qy         61 CAA--------------------AGCGATTTAGAACAAGAT----------------- 81
              :::                    ||||||:::|||
Db        966 LysValThrThrGluAlaLysLeuLysLysLeuGluGluAspValIleValLeuGluAsp 985

Qy         82 ---------AGACTTGCTAAAGAAAAG------TTACAAGAGCAGCAAAGCGATTTA--- 123
                     :::|||||||||||||||      |||:::::::::   |||:::
Db        986 GlnAsnLeuLysLeuAlaLysGluLysLysLeuGluAspArgMetSerGluPheThr 1005

Qy        124 ---------------GAACAAGAGAGA---------CTTGCTAAAGAAAAGTTGCAAGAA 159
                          |||:::|||:::         |||||||||   |||   :::
Db       1006 ThrAsnLeuThrGluGluGluGluLysSerLysSerLeuAlaLysLeuLysAsnLysHis 1025

Qy        160 CAACAA------AGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 213
              :::         :::|||||||||:::      |||   :::||||||         |||
Db       1026 GluAlaMetIleThrAspLeuGluGluArgLeuArgArgGluGluLys---------Gln 1042

Qy        214 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273
              :::   :::||||||:::   ||||||         ||||||:::         ||||||
Db       1043 ArgGlnGluLeuGluLysThrArgArg---------LysLeuGluGlyAspSerSerAsp 1059

Qy        274 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAA--- 330
              |||      |||::::::|||      :::||||||   |||   ::::::|||:::
Db       1060 Leu---HisAspGlnIleAla------GluLeuGlnAlaGlnIleAlaGluLeuLysIle 1076

Qy        331 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG--- 387
              |||      :::   :::|||:::||||||      :::   :::|||:::|||
Db       1077 GlnLeuSerLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgValGluGluGluAla 1096

Qy        388 ----------------------------AGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
                                         |||      :::   ::::::   |||
Db       1097 AlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGluSerGlnIleThrGluLeu 1116

Qy        418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
              |||   |||||||||   ||||||   ::::::   |||   ::::::|||:::   |||
Db       1117 GlnGluAspLeuGluSerGluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAsp 1136

Qy        478 TTAGAACAAGAGAGACGTGCT---AAAGAAAAGTTG---------------------- 510
              |||   :::|||      |||   |||   :::|||
Db       1137 LeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAla 1156

Qy        511 ------CAAGAACAACAAAGCGATTTAGAACAAGAG---------------------- 540
                   ||||||   :::|||      |||||||||
Db       1157 AlaGlnGlnGluLeuArgSerLysArgGluGlnGluValThrValLeuLysLysThrLeu 1176

Qy        541 ---------------------------------------------AGACGTGCTAAAGAA 555
                                                          ::::::      :::
Db       1177 GluAspGluAlaLysThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGln 1196

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 615
              ::::::|||                |||||||||   :::|||   |||      |||:::
Db       1197 AlaIleGluGluLeuAlaGluGlnLeuGluGlnThrLysArgValLysAlaAsnLeuGlu 1216

Qy        616 GAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACG---------------------- 654
              :::   ::::::   ||||||      :::|||:::
Db       1217 LysAlaLysGlnAlaLeuGluSerGluArgAlaGluLeuSerAsnGluValLysValLeu 1236

Qy        655 ---------AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGAT 705
                      |||   :::   |||   |||::::::      |||         ::::::
Db       1237 LeuGlnGlyLysGlyAspAlaGluHisLysArgLysLysValAspAlaGlnLeuGlnGlu 1256

Qy        706 TTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATAT 765
              |||      :::                        :::|||   ||||||
Db       1257 LeuGlnValLysPhe----------------------ThrGluGlyGluArg------ 1266

Qy        766 TATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAG 825

Db       1267 ----------------------------------------------------------Val 1267

Qy        826 GAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATT---ACAACAAATGTTGAAGGA 882
              :::      :::   |||:::   |||:::      :::         ||||||   |||
Db       1268 LysThrGluLeuAlaGluArgValAsnLysLeuGlnValGluLeuAspAsnValThrGly 1287

Qy        883 ---CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                        :::::      ||||||   |||
Db       1288 LeuLeuAsnGlnSerAspSerLysSerIleLysLeuAlaLysAspPheSerAlaLeuGlu 1307

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||   :::|||      :::|||      |||         ::::::
Db       1308 SerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThrArgLeuLysLeuSerPhe 1327

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||                :::::   :::   |||||||||   :::|||
Db       1328 SerThr------------LysLeuLysGlnThrGluAspGluLysAsnAla---LeuLys 1342

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              :::   |||   :::|||||||||         :::|||::::::         ::::::
Db       1343 GluGlnLeuGluGluGluGluGluAlaLysArgAsnLeuGluLysGlnIleSerValLeu 1362

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||                    :::::   :::         |||:::::::::|||
Db       1363 Gln------------------GlnGlnAlaValGluAlaArgLysLysMetAspAspGly 1376

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              :::         |||      ::::::      :::   ::::::|||      |||
Db       1377 LeuGlyCysLeuGluIleAlaGluGluAlaLysLysLysLeuGlnLysAspLeuGluSer 1396

Qy       1240 TTATCT---------GAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGAT 1290
              |||:::         ||||||||||||         |||   ||||||||||||
Db       1397 LeuThrGlnArgTyrGluGluLysIleAlaAlaTyrAspLysLeuGluLysThrLysThr 1416

Qy       1290 ---------------------------------------------------------- 1290

Db       1417 ArgLeuGlnGlnGluLeuAspAspIleAlaValAspLeuAspHisGlnArgGlnThrVal 1436

Qy       1291 AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAAT 1350
              :::|||   :::         |||      :::|||:::   :::   :::   ||||||
Db       1437 SerAsnLeuGluLysLysGlnLysLys---PheAspGlnLeuLeuAlaGluGluLysAsn 1455

Qy       1351 --------------------------GATAAGCAGGTTAATAAGGAAAAGGAAAAATTC 1383
                                       |||:::         :::   :::||||||
Db       1456 IleSerAlaLysTyrAlaGluGluArgAspArgAlaGluAlaGluAlaArgGluLysGlu 1475

Qy       1384 ATAAAATCATTG 1395
                 |||:::|||
Db       1476 ThrLysAlaLeu 1479

RESULT 28
MWKW
myosin heavy chain B [similarity] - Caenorhabditis elegans
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Caenorhabditis elegans
C;Date: 13-Jun-1983 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: T20770; T21629; A93958; A93287; A21074; A02992
R;Kershaw, J.
```

```
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19322
A;Accession: T20770
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1963 <WIL>
A;Cross-references: UNIPROT:O02244; UNIPARC:UPI0000061004; EMBL:Z81499; PIDN:CAB04089.1;
A;Experimental source: clone F11C3
A;Accession: T21629
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1963 <WI2>
A;Cross-references: UNIPARC:UPI0000061004; EMBL:Z83107; PIDN:CAB05505.1; GSPDB:GN00019;
A;Experimental source: clone F32A7
R;Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A;Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy ch
A;Reference number: A93958; MUID:83273600; PMID:6576334
A;Accession: A93958
A;Molecule type: DNA
A;Residues: 1-61,'EMSVIQ',65-376,'V',378-1963 <KAR>
A;Cross-references: UNIPARC:UPI000012FBDC; GB:J01050; NID:g156399; PIDN:AAA28124.1; PID:
R;McLachlan, A.D.; Karn, J.
Nature 299, 226-231, 1982
A;Title: Periodic charge distributions in the myosin rod amino acid sequence match cross
A;Reference number: A93287; MUID:82272395; PMID:7202124
A;Accession: A93287
A;Molecule type: DNA
A;Residues: 847-1333,'R',1335-1876,'L',1878-1963 <MCL>
A;Cross-references: UNIPARC:UPI0000173E26
R;Wills, N.; Gesteland, R.F.; Karn, J.; Barnett, L.; Bolten, S.; Waterston, R.H.
Cell 33, 575-583, 1983
A;Title: The genes sup-7 X and sup-5 III of Caenorhabditis elegans suppress amber nonsen
A;Reference number: A21074; MUID:83232892; PMID:6571695
A;Accession: A21074
A;Molecule type: DNA
A;Residues: 1873-1963 <WI3>
A;Cross-references: UNIPARC:UPI000016B912; GB:V01494; GB:J01049; NID:g6783; PIDN:CAA2473
C;Genetics:
A;Gene: unc-54; CESP:F11C3.3
A;Map position: 1
A;Introns: 21/3; 64/3; 111/3; 264/1; 525/3; 951/2; 1747/3; 1819/3; 1894/3
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co
F;84-775/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;662-684/Region: actin binding #status predicted
F;766-780/Region: actin binding #status predicted
F;848-1963/Domain: coiled coil #status predicted <COI>
F;848-1162/Region: S2
F;1163-1963/Region: light meromyosin
F;125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;702,712/Active site: Cys #status predicted

Alignment Scores:
Pred. No.:              1.36e-09      Length:         1963
Score:                  310.00        Matches:        151
Percent Similarity:     37.22%        Conservative:   114
Best Local Similarity:  21.21%        Mismatches:     185
Query Match:            12.46%        Indels:         262
DB:                     1             Gaps:           27

US-09-837-344-45 (1-1482) x MWKW (1-1963)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::|||      |||||||||          |||      |||         |||:::
Db        960 LysLysGlnIleGlnAspLeuGluMetSerLeuArgLysAlaGlu------SerGluLys 977

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||          |||   |||      ::::::::::::|||:::|||
Db        978 GlnSerLysAspHisGlnIleArgSerLeuGlnAspGluMetGlnGlnGlnAspGluAla 997

Qy        121 TTA----------------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 156
              :::                      ||||||   :::   :::::::   ::::::
Db        998 IleAlaLysLeuAsnLysGluLysLysHisGlnGluGluIleAsnArgLysLeuMetGlu 1017

Qy        157 GAACAACAAAGCGATCTAGAACAA---------GAGAGACGTGCTAAAGAAAAGTTGCAA 207
              :::   ||||||:::   ::::::         :::   :::   |||   ||||||:::
Db       1018 AspLeuGlnSerGluGluAspLysGlyAsnHisGlnAsnLysValLysAlaLysLeuGlu 1037

Qy        208 GAACAACAAAGCGAT------------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
              :::          |||          ||||||:::|||:::||||||:::      |||
Db       1038 GlnThrLeuAspAspLeuGluAspSerLeuGluArgGluLysArgAlaArgAlaAspLeu 1057

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAG 312
              :::|||:::      :::|||   :::   ::::::|||:::|||   :::   |||
Db       1058 AspLysGlnLysArgLysValGluGlyGluLeuLysIleAlaGlnGluAsnIleAspGlu 1077

Qy        313 ---------CAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA---AAGTTGCAA 360
                     |||:::   ||||||||          :::   ||||||   :::|||
Db       1078 SerGlyArgGlnArgHisAspLeuGluAsnAsnLeuLysLysLysGluSerGluLeuHis 1097

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
                   |||   ||||||   |||:::            |||||||||   |||
Db       1098 SerValSerSerArgLeuGluAspGluGlnAlaLeuValSerLysLeuGlnArgGlnIle 1117

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              |||                  :::   ::::::   |||   :::   :::|||||||
Db       1118 LysAsp-----------------GlyGlnSerArgIleSerGluLeuGluGluGluLeu 1131

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              |||   |||||:::::::::   |||           :::|||||||||:::::|||
Db       1132 GluAsnGluArgGlnSerArgSerLysAlaAspArgAlaLysSerAspLeuGlnArgGlu 1151

Qy        541 ---AGACGTGCTAAAGAAAAGTTGCAAGAACAA------------CAAAGCGATTTAGAA 585
                    |||||||||   ||||||                 :::   :::|||
Db       1152 LeuGluGluLeuGlyGluLysLeuAspGluGlnGlyGlyAlaThrAlaAlaGlnValGlu 1171

Qy        586 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA------ 639
                  ::::::   :::   :::|||   :::   :::|||||||||||:::
Db       1172 ValAsnLysLysArgGluAlaGluLeuAlaLysLeuArgArgAspLeuGluGluAlaAsn 1191

Qy        640 ------------------------------AGGAAGGCTGATACG-------------- 654
                                           ||||||     |||
Db       1192 MetAsnHisGluAsnGlnLeuGlyGlyLeuArgLysLysHisThrAspAlaValAlaGlu 1211

Qy        655 --------------------------------AAAAAAAATTTAGAAAGAAAAAAGGAA 681
                                             |||      :::|||:::   |||
Db       1212 LeuThrAspGlnLeuAspGlnLeuAsnLysAlaLysAlaLysValGluLysAspLysAla 1231

Qy        682 CATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA---------------------- 720
                          ||||||||||||      :::|||
Db       1232 GlnAlaValArgAspAlaGluAspLeuAlaAlaGlnLeuAspGlnGluThrSerGlyLys 1251

Qy        721 ------------------------------GAAATACCAGCTATAGAACTTCCATCAGAA 750
                                           .|||:::          ||||||   |||:::
Db       1252 LeuAsnAsnGluLysLeuAlaLysGlnPheGluLeuGlnLeuThrGluLeuGlnSerLys 1271

Qy        751 ---AATGAACGTGGATATTATATACCACATCAATCTTCTTTA------------------ 789
                 :::|||:::         :::         :::||||||
Db       1272 AlaAspGluGlnSerArgGlnLeuGlnAspPheThrSerLeuLysGlyArgLeuHisSer 1291

Qy        790 ------------------------------------CCTCAGGACAACAGAGGGAAT 810
                                                   |||   |||:::
Db       1292 GluAsnGlyAspLeuValArgGlnLeuGluAspAlaGluSerGlnValAsnGlnLeuThr 1311

Qy        811 AGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA--------------- 855
                  |||:::     |||   :::|||:::     |||
Db       1312 ArgLeuLysSerGlnLeuThrSerGlnLeuGluGluAlaArgArgThrAlaAspGluGlu 1331

Qy        855 ------------------------------------------------------------ 855

Db       1332 AlaArgGluArgGlnThrValAlaAlaGlnAlaLysAsnTyrGlnHisGluAlaGluGln 1351

Qy        856 ------GAATCTATTACAACAAATGTTGAAGGACGAAGG---------------------- 888
```

```
                  ||||||:::         :::||||||:::
Db       1352 LeuGlnGluSerLeuGluGluIleGluGlyLysAsnGluIleLeuArgGlnLeuSer 1371

Qy        889 ------------GATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA 936
                          ||||||   :::               :::   :::|||   :::|||
Db       1372 LysAlaAsnAlaAspIleGlnGlnTrpLysAlaArgPheGluGlyGluGlyLeuLeuLys 1391

Qy        937 CCAGAACAAAAAGAAGAT---------------------------------------- 954
                 ::::::   ||||||
Db       1392 AlaAspGluLeuGluAspAlaLysArgArgGlnAlaGlnLysIleAsnGluLeuGlnGlu 1411

Qy        955 ------AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAAT---------ATT 999
                  :::|||:::   :::|||   :::||||||                      :::
Db       1412 AlaLeuAspAlaAlaAsnSerLysAsnAlaSerLeuGluLysThrLysSerArgLeuVal 1431

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGT------------------------------- 1026
                |||::::::|||   |||:::
Db       1432 GlyAspLeuAspAspAlaGlnValAspValGluArgAlaAsnGlyValAlaSerAlaLeu 1451

Qy       1027 ----------------------------------------------AAGTATGAGGATGAA 1041
                                                           |||     |||:::
Db       1452 GluLysLysGlnLysGlyPheAspLysIleIleAspGluTrpArgLysLysThrAspAsp 1471

Qy       1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
                ::::::||||||   |||   :::      |||                |||||
Db       1472 LeuAlaAlaGluLeuAspGlyAlaGlnArgAspLeuArgAsnThrSerThrAspLeuPhe 1491

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAA----------------- 1143
                :::   |||                  |||   |||:::|||
Db       1492 LysAlaLys-----------------AsnAlaGlnGluGluLeuAlaGluValValGlu 1505

Qy       1144 ------------GAAAACATAGGAATTTATAAAGAACTAGAAGATTTG------------ 1179
                         ||||||      :::   :::|||::::::||||||
Db       1506 GlyLeuArgArgGluAsnLysSerLeuSerGlnGluIleLysAspLeuThrAspGlnLeu 1525

Qy       1180 ----------------------------------------------------ATAGAG 1185
                                                                   ||||||
Db       1526 GlyGluGlyGlyArgSerValHisGluMetGlnLysIleIleArgArgLeuGluIleGlu 1545

Qy       1186 AAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCT 1245
                |||   |||             |||||||||          :::::::   |||
Db       1546 LysGluGluLeuGlnHisAlaLeuAspGluAla------GluAlaAlaLeuGluAlaGlu 1563

Qy       1246 GAAGAAAAAATA---------------------------------------------- 1257
                |||   |||:::
Db       1564 GluSerLysValLeuArgAlaGlnValGluValSerGlnIleArgSerGluIleGluLys 1583

Qy       1258 ---------AAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                      :::|||   :::::::::|||   |||:::   |||:::   :::
Db       1584 ArgIleGlnGluLysGluGluGluPheGluAsnThrArgLysAsnHisAlaArgAlaLeu 1603

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                :::                                              |||   :::
Db       1604 Glu-------------------------------------SerMetGlnAlaSerLeu 1610

Qy       1369 GAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTA 1428
                |||   |||                          |||      |||:::|||
Db       1611 GluThrGluAla--------------------------LysGlyLysAlaGluLeuLeu 1621

Qy       1429 CAGATCGTGGATGAGTTATCTGAAGATATAACTAAA 1464
                :::|||      :::|||    ||||||   :::
Db       1622 ArgIleLysLysLysLeuGluGlyAspIleAsnGlu 1633

RESULT 29
T16416
hypothetical protein F52B10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T16416
R;Wilcox, L.
submitted to the EMBL Data Library, December 1995
A;Description: The sequence of C. elegans cosmid F52B10.
A;Reference number: Z18510
A;Accession: T16416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1956 <WIL>
A;Cross-references: UNIPROT:Q20641; UNIPARC:UPI0000177613; EMBL:U41990; NID:g1118154; P
C;Genetics:
A;Gene: CESP:F52B10.1
A;Introns: 44/2; 108/3; 161/1; 211/3; 242/3; 297/1; 345/1; 503/1; 876/3; 941/3; 1134/2;
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;81-763/Domain: myosin motor domain homology <MMO>

Alignment Scores:
Pred. No.:                1.45e-09      Length:          1956
Score:                    309.50        Matches:         142
Percent Similarity:       41.06%        Conservative:    122
Best Local Similarity:    22.08%        Mismatches:      204
Query Match:              12.44%        Indels:          175
DB:                       2             Gaps:            26

US-09-837-344-45 (1-1482) x T16416 (1-1956)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ::::::   :::       ||||||||||||   :::|||::::::||||||
Db        945 ArgAspLeuGluGluGlnLeuGluGluGlnGluGluGlnAlaArgGlnLysLeuLeuAsp 964

Qy         61 CAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 117
              ::::::::::::::|||       |||      :::|||:::|||   |||   |||
Db        965 LysThrAsnValAspGlnArgLeuArgAsnLeuGluGluArgLeuValGluLeuGlnAsp 984

Qy        118 GAT------------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAA 165
                             |||:::::::::||||||   :::||||||::::::
Db        985 AlaTyrAspLysLeuLeuLysGluLysArgLeuLeuGluGluLysValGluGlyLeuThr 1004

Qy        166 AGCGAT---CTAGAACAAGAGAGACGTGCTAAA-------------------------- 195
              :::       |||:::   |||   |||||||||
Db       1005 ThrGlnLeuLeuAspHisGluGluArgAlaLysHisGlyValLysAlaLysGlyArgLeu 1024

Qy        196 ---GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 252
                 :::|||   |||   :::   ||||||   :::|||||:::   |||   :::
Db       1025 GluAsnGlnLeuHisGluLeuGluGlnAspLeuAsnArgGluArgGlnTyrLysSerGlu 1044

Qy        253 TTGCAAGAACAACAAAGCGATTTA-----------------GAACAAGATAGACTTGCT 294
              |||::::::   :::       |||                        :::|||   ||||||
Db       1045 LeuGluGlnHisLysArgLysLeuLeuAlaGluLeuGluAspSerLysAspHisLeuAla 1064

Qy        295 AAA-------------------------------------------------GAAAAGTTA 306
              :::                                                 |||:::|||
Db       1065 GluLysMetGlyLysValGluGluLeuAsnAsnGlnLeuMetLysArgAspGluGluLeu 1084

Qy        307 CAAGAGCAG------------------------------------------------- 315
              |||   |||
Db       1085 GlnHisGlnLeuThrArgTyrAspGluGluSerAlaAsnValThrLeuMetGlnLysGln 1104

Qy        316 --------------------------------CAAAGCGATTTAGAACAAGAGAGACGT 342
                                             :::   |||:::|||   ||||||
Db       1105 MetArgAspMetGlnThrThrIleAspGluLeuArgGluAspMetGluThrGluArgAsn 1124

Qy        343 GCTAAAGAAAAG----------------------------------------------- 354
              |||:::   |||
Db       1125 AlaArgAsnLysAlaGluMetThrArgArgGluValValAlaGlnLeuGluLysValLys 1144

Qy        355 ----------------------------------TTGCAAGAACAACAAAGCGATTTA 378
                                               ||||||:::      |||
Db       1145 GlyAspValLeuAspLysValAspGluAlaThrMetLeuGlnAspLeuMetSerArgLys 1164

Qy        379 GAACAAGAG------------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 426
              :::::::|||            :::||||||   |||::::::|||
Db       1165 AspGluGluValAsnAlaThrLysArgAlaIleGluGlnIleGlnHisThrMetGluGly 1184

Qy        427 TTAGAACAAGAGAGACTTGCTAAA------GAAAAGTTGCAAGAACAACAAAGCGATTTA 480
                 :::|||:::   ||||||         :::::::::|||          :::
Db       1185 LysIleGluGluGlnLysAlaLysPheSerArgGlnValGluGluLeuHisAspGlnIle 1204
```

```
Qy       481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
Db      1205 GluGlnHisLysLysGlnArgSerGlnLeuGluLysGlnGlnAsnGlnAlaAspGlnGlu 1224

Qy       541 AGA------CGTGCTAAAGAA------AAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db      1225 ArgAlaAspMetAlaGlnGluIleAlaLeuLeuGlnAlaSerArgAlaAspIleAspLys 1244

Qy       589 GAGAGACGTGCTAAAGAA---AAGTTGCAAGAGCAGCAAAGAGATTTAGAA--------- 636
Db      1245 LysArgLysIleHisGluAlaHisLeuMetGluIleGlnAlaAsnLeuAlaGluSerAsp 1264

Qy       637 ----------------------------CAAAGGAAGGCTGATACGAAAAAAAATTTA 666
Db      1265 GluHisLysArgThrLeuIleAspGlnLeuGluArgSerArgAspGluLeuAspHisLeu 1284

Qy       667 GAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATA 726
Db      1285 AsnArgValArgGluGluGluGluHisAlaPheAlaAsnMetGlnArgArgLeuAlaThr 1304

Qy       727 CCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCT 786
Db      1305 AlaGluGlyGlnIleGlnGluLeuAsnGluGln---------IleGlnGluGluThrArg 1321

Qy       787 TTACCTCAGGACAACAGAGGGAATAGTAGA---------GATTCCAAGGAAATGTCTATA 837
Db      1322 LeuLysIleAlaAsnIleAsnArgAlaArgGlnLeuGluAspGluLysAsnAlaLeuLeu 1341

Qy       838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db      1342 AspGluLysGluGluAlaGluGlyLeuArgAlaHisLeuGlu------LysGluIleHis 1359

Qy       898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db      1360 AlaAla--------------ArgGlnGlyAlaGlyGluAlaArgArgLysAlaGluGlu 1374

Qy       958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
Db      1375 SerValAsnGlnGlnLeuGluGluLeuArgLysLysAsnLeuArgAspValGluHisLeu 1394

Qy      1018 CAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA---------TATGACGATTCATTA 1068
Db      1395 GlnLysGlnLeuGluGluSerGluValAlaLysGluArgIleLeuGlnSerLysLysLys 1414

Qy      1069 ATAGATGAAGAAGAAGATGATGAAGACTTA---------------------------- 1098
Db      1415 IleGlnGlnGluLeuGluAspSerSerMetGluLeuGluAsnValArgAlaSerHisArg 1434

Qy      1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
Db      1435 AspSerGluLysArgGlnLysLysPheGluSerGlnMetAlaGluGluArgValAlaVal 1454

Qy      1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGAT------------ 1206
Db      1455 GlnLysAlaLeuLeuAspArgAspAlaMetSerGlnGluLeuArgAspArgGluThrArg 1474

Qy      1207 ------------TTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAA 1254
Db      1475 ValLeuSerLeuLeuAsnGlu---ValAspIleMetLysGluHisLeuGluGluSerAsp 1493

Qy      1255 ATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAA 1314
Db      1494 ArgValArgArgSerLeuGlnGlnGluLeuGlnAspSerIleSerAsnLysAspAspPhe 1513

Qy      1315 AGTTTGTATGATGAGCATATTAAAAAATATAAA----------------------AATGAT 1353
Db      1514 GlyLysAsnValHisGluLeuGluLysAlaLysArgSerLeuGluAlaGluLeuAsnAsp 1533

Qy      1354 AAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGA 1413
Db      1534 MetArgValGlnMetGluGluLeuGlu------------------------------- 1542

Qy      1414 GACAATGAAATTTTACAGATCGTGGATGAG------TTATCTGAAGATATAACTAAATAT 1467
```

```
Db      1543 AspAsn------LeuGlnIleAlaGluAspAlaArgLeuArgLeuGluValThrAsnGln 1560

Qy      1468 TTTATGAAA 1476
Db      1561 AlaLeuLys 1563
```

RESULT 30
A41604
myosin heavy chain, smooth muscle, long splice form - rabbit
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C;Accession: A41604; A33501
R;Babij, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A;Title: Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete
A;Reference number: A41604; MUID:92073350; PMID:1961735
A;Accession: A41604
A;Molecule type: mRNA
A;Residues: 1-1972 <BAB>
A;Cross-references: UNIPARC:UPI0000173E2A; GB:M77812
R;Nagai, R.; Kuro-o, M.; Babij, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A;Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
A;Reference number: A33501; MUID:89255535; PMID:2722872
A;Accession: A33501
A;Molecule type: mRNA
A;Residues: 1455-1972 <NAG>
A;Cross-references: UNIPARC:UPI0000000396; GB:J04833; NID:g165519; PIDN:AAA31407.1; PID:
A;Experimental source: smooth muscle
A;Note: examination by Southern blotting for the regions of difference between this isof
ative splicing
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;88-771/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;559-572/Region: actin binding #status predicted
F;633-647/Region: actin binding #status predicted
F;844-1938/Domain: coiled coil #status predicted <COI>
F;844-1284/Region: S2
F;1285-1972/Region: light meromyosin
F;1939-1972/Domain: carboxyl-terminal <CBT>
F;129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;701,711/Active site: Cys #status predicted

Alignment Scores:

Pred. No.:	1.54e-09	Length:	1972
Score:	309.00	Matches:	150
Percent Similarity:	41.95%	Conservative:	113
Best Local Similarity:	23.92%	Mismatches:	162
Query Match:	12.42%	Indels:	202
DB:	1	Gaps:	30

US-09-837-344-45 (1-1482) x A41604 (1-1972)

```
Qy         4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
Db       932 GluGluGluGluAspArgGlyGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGln 951

Qy        64 ---------------AGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAG 108
Db       952 MetLeuAspLeuGluGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeu 971

Qy       109 CAGCAA-----------------------------------------------AGCGAT 120
Db       972 GluLysValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAsp 991

Qy       121 TTAGAACAAGAGAGACTTGCTAAAGAAAAG------TTGCAAGAACAACAAAGCGATCTA 174
Db       992 AspGlnAsnAsnLysLeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeu 1011
```

```
Qy       175 --------------GAACAAGAGAGACGTGCTAAAGAA-------------------- 198
                           |||:::|||   :::||||||
Db      1012 ThrThrAsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysAsnLys 1031

Qy       199 ----------------------------AAGTTGCAAGAACAACAAAGCGATTTAGAA 228
                                        :::|||:::::::::::::         :::
Db      1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGln 1051

Qy       229 CAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT--- 285
             :::      :::   |||   |||:::      :::   |||||||||   :::
Db      1052 GluLeuGluLysLeuLysArgLysMetAspGlyGluAlaSerAspLeuHisGluGlnIle 1071

Qy       286 ----------------------------------AGACTTGCTAAA------GAAAAG 303
                                               :::|||||||||      |||:::
Db      1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu 1091

Qy       304 TTACAAGAGCAGCAAAGCGATTTAGAA------------CAAGAGAGACGTGCTAAAGAA 351
             ||||||         :::  ||||||           |||:::     |||   :::
Db      1092 LeuGlnAlaAlaLeuAlaArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLys 1111

Qy       352 AAGTTGCAAGAA--------------------CAACAAAGCGATTTAGAACAAGAGAGA 390
             |||::::::|||                    |||   ||||||:::   ||||||
Db      1112 LysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluArg 1131

Qy       391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCT--- 447
                |||:::   |||   ::::::|||:::   ||||||   :::|||      |||
Db      1132 AlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluLeuGluAlaLeu 1151

Qy       448 AAAGAAAAGTTGCAAGAA------CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 501
             |||   :::|||::::::          :::      :::||||||   |||||||||
Db      1152 LysThrGluLeuGluAspThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLys 1171

Qy       502 GAAAAG--------------------------TTGCAAGAACAACAAAGCGATTTAGAA 534
                 :::                         |||   |||:::             |||
Db      1172 ArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGlu 1191

Qy       535 ---------CAAGAGAGACGTGCTAAA--------------GAAAAGTTGCAAGAACAA 570
                     ||||||   |||   |||                |||:::|||   ||||||
Db      1192 AlaGlnValGlnGluMetArgGlnLysHisThrGlnValValGluGluLeuThrGluGln 1211

Qy       571 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGAT 630
                 |||||||||   :::||||||||||      |||   :::   ::::::
Db      1212 ---------LeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysThrLysGlnThr 1228

Qy       631 TTAGAACAAAGGAAGGCTGAT----------------------------ACGAAAAAA 660
             ||||||:::      ||||||                            |||:::
Db      1229 LeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGlyGlnAlaLysGln 1248

Qy       661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
                :::|||   ||||||:::                                 :::|||
Db      1249 GluValGluHisLysLysLys-------------------------------LysLeu 1257

Qy       721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
             |||:::        ||||||   |||:::         |||
Db      1258 GluValGlnLeuGlnGluLeuGlnSerLysCysSerAspGly---------------- 1271

Qy       781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
                                         |||         |||::::::
Db      1272 ----------------------------GluArgAlaArgAlaGluLeuAsn------ 1279

Qy       841 GAAAAAACAAATAGA--------------GAATCTATTACAACAAATGTTGAAGGACGA 885
             :::|||   ::::::              ||||||:::|||
Db      1280 AspLysValHisLysLeuGlnAsnGluValGluSerValThr---------------- 1293

Qy       886 AGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAA--------- 936
                         |||   |||   |||   :::      :::||||||
Db      1294 --------------GlyMetLeuSerGluAlaGluGlyLysAlaIleLysLeuAlaLys 1308

Qy       937 --------------CCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACA 981
                           |||   :::|||             :::|||      |||
Db      1309 GluValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluGluThr 1328

Qy       982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAA 1041
```

```
                :::   :::|||:::|||                :::::   :::   |||||||||
Db      1329 ArgGlnLysLeuAsnValSerThr------------LysLeuArgGlnLeuGluAspGlu 1344

Qy      1042 ATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC 1101
                                  :::||||||      :::|||:::   ||||||:::::::::
Db      1345 --------------ArgAsnSerLeu------GlnGluGlnLeuAspGluGluMetGlu 1357

Qy      1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
                  :::   :::   ::::::   :::       :::   :::
Db      1358 AlaLysGlnAsnLeuGluArgHisIleSerThrLeuAsnIleGlnLeuSerAspSerLys 1377

Qy      1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATA 1221
             |||:::|||:::|||                |||:::|||::::::      :::
Db      1378 LysLysLeuGlnAspPheAlaSerThrValGluSerLeuGluGluGlyLysLysArgPhe 1397

Qy      1222 GAAAAATCATCAGAAGAATTATCT---------GAAGAAAAAATAAAAAAAGGAAAGAAA 1272
             :::|||      |||  |||:::         |||||||||                |||
Db      1398 GlnLysGluIleGluSerLeuThrGlnGlnTyrGluGluLysAlaAlaAlaTyrAspLys 1417

Qy      1273 TATGAAAAAACAAAGGAT---------------------------------------- 1290
                |||||||||||||:::
Db      1418 LeuGluLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAsp 1437

Qy      1291 -----------------AATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCAT 1332
                              :::|||   :::          |||      :::|||:::
Db      1438 AsnGlnArgGlnLeuValSerAsnLeuGluLysLysGlnLysLys---PheAspGlnLeu 1456

Qy      1333 ATTAAAAAATATAAAAAT--------------------------GATAAGCAGGTTAAT 1365
             :::   :::  ||||||                          |||:::
Db      1457 LeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArgAspArgAlaGluAla 1476

Qy      1366 AAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
             :::   :::||||||      |||:::|||      :::               |||
Db      1477 GluAlaArgGluLysGluThrLysAlaLeu---SerLeuAlaArgAlaLeuGluGluAla 1495

Qy      1426 TTACAGATCGTGGATGAGTTA 1446
             |||:::      :::||||||
Db      1496 LeuGluAlaLysGluGluLeu 1502
```

Search completed: December 21, 2005, 11:33:17
Job time : 191.302 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:09:34 ; Search time 7876.23 Seconds
(without alignments)
8803.514 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 1482
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	289.8	19.6	317	10	T09652	T09652 0224m3 gmbP
	2	254.4	17.2	1721	8	DR145459	DR145459 49056454
	3	251	16.9	1808	8	DR125630	DR125630 49096520
c	4	246.8	16.7	1519	10	AG386893	AG386893 Mus muscu
c	5	245.6	16.6	1453	10	AJ591978	AJ591978 Arabidops
	6	242.8	16.4	2071	8	DR147304	DR147304 49027161
	7	241.4	16.3	1932	8	DR141720	DR141720 49300102
c	8	240	16.2	1491	10	CG753221	CG753221 P048-2-A0
	9	239.4	16.2	1727	10	CL114085	CL114085 ISB1-59J1
	10	236.8	16.0	1783	8	DR141924	DR141924 49190649
c	11	236.2	15.9	1434	10	AJ592058	AJ592058 Arabidops
	12	236.2	15.9	1808	10	CL078963	CL078963 CH216-154
	13	235.8	15.9	1797	10	CL066150	CL066150 CH216-108
	14	233.4	15.7	1654	10	CL118709	CL118709 ISB1-72I1
c	15	232.6	15.7	1811	10	CG753732	CG753732 P048-4-G0
	16	232.4	15.7	1829	10	CL109409	CL109409 ISB1-51H2
	17	232.2	15.7	1736	10	CL066373	CL066373 CH216-108
c	18	232	15.7	1407	10	AJ592026	AJ592026 Arabidops
	19	232	15.7	1753	8	DR131208	DR131208 49217594
	20	231.2	15.6	1512	10	CL113943	CL113943 ISB1-59E2
	21	231.2	15.6	1531	10	CG748014	CG748014 P041-4-B0
	22	230	15.5	1917	10	CL081899	CL081899 CH216-164
	23	229.8	15.5	1650	10	CL041163	CL041163 CH216-51P
	24	229.6	15.5	1762	10	CL082587	CL082587 CH216-168
	25	228.8	15.4	1824	10	CL081234	CL081234 CH216-161
	26	228.6	15.4	1727	10	CL118815	CL118815 ISB1-72O5
c	27	228.4	15.4	1507	10	AG346189	AG346189 Mus muscu
	28	228.4	15.4	1632	10	CL082569	CL082569 CH216-167
	29	228.2	15.4	1616	10	CL081995	CL081995 CH216-165
	30	228	15.4	1630	10	CL078686	CL078686 CH216-152
	31	227.6	15.4	1457	10	CL082658	CL082658 CH216-169
c	32	227.2	15.3	1566	10	CG757757	CG757757 P053-1-D0
	33	227	15.3	1434	9	CC187638	CC187638 CH261-98P
	34	226.8	15.3	1823	10	CL058635	CL058635 CH216-88H
	35	226.4	15.3	1259	11	CR004722	CR004722 Forward s
	36	226.4	15.3	2033	10	CL081977	CL081977 CH216-165
c	37	226.4	15.3	2103	10	CL041093	CL041093 CH216-51M
	38	226	15.2	1712	10	CL078539	CL078539 CH216-151
	39	225.4	15.2	1938	8	DR142277	DR142277 49191190
	40	225.2	15.2	1547	10	CL066420	CL066420 CH216-108
	41	225	15.2	1377	10	CL077307	CL077307 CH216-144
	42	224.8	15.2	1716	9	CC222065	CC222065 CH261-11A
	43	224	15.1	1562	10	CL078533	CL078533 CH216-151
	44	224	15.1	1651	10	CL079030	CL079030 CH216-154
c	45	223.8	15.1	1482	10	AG332292	AG332292 Mus muscu

ALIGNMENTS

RESULT 1
T09652
LOCUS T09652 317 bp DNA linear GSS 05-DEC-2000
DEFINITION 0224m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic clone 0224m, genomic survey sequence.
ACCESSION T09652
VERSION T09652.1 GI:319484
KEYWORDS GSS.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 317)
AUTHORS Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C. and Dame,J.B.
TITLE Gene sequence tags from Plasmodium falciparum genomic DNA fragments prepared by the genease activity of mung bean nuclease
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
PUBMED 8234327
COMMENT Other_GSSs: 0224m7
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .317
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/db_xref="taxon:5833"
/clone="0224m"
/lab_host="E. coli XL1-Blue"
/clone_lib="gmbPfHB3.1, G. Roman Reddy"
/note="Vector: pBlueScript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were polished using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBlueScript SK(+). Recombinant plasmids

transformed E. coli XL1-Blue."
ORIGIN

Query Match 19.6%; Score 289.8; DB 10; Length 317;
Best Local Similarity 96.2%; Pred. No. 6.6e-38;
Matches 305; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```
Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||| |   |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 GAATTCGAAGCTATAGAACTTCCATCAGAGAATGAACGTGGATATTATATACCACATCAA 60

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         61 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATATCTATAATA 120

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 180

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        181 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGTACAAAAAGAAGATAAATCT 240

Qy        961 GCTGACATACAAAATCATACATTAGAGA-CAGTAAATATTTCTGATGTTAATGATTTTCA 1019
              |||||||||||||||||||||| |||  |||||| ||||||| ||||||||||||||||
Db        241 GCTGACATACAAAATCATACATTNGAGNCCAGTAANTATTTCTNATGTTAATGATTTTCA 300

Qy       1020 AATAAGTAAGTATGAGG 1036
              |||||||||||||||||
Db        301 AATAAGTAAGTATGAGG 317
```

RESULT 2
DR145459
LOCUS DR145459 1721 bp mRNA linear EST 16-JUN-2005
DEFINITION 49056454 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone A9 3', mRNA sequence.
ACCESSION DR145459
VERSION DR145459.1 GI:67890549
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1721)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226693903
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1721
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="A9"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

Query Match 17.2%; Score 254.4; DB 8; Length 1721;
Best Local Similarity 47.6%; Pred. No. 4.2e-32;
Matches 660; Conservative 0; Mismatches 728; Indels 0; Gaps 0;

```
Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
              || || |  ||| |   | |  || | | |     ||| ||||    || || ||
Db        137 AAAAAAANAAAAACCCAAAAAAGAAAAAAANANNAAAAAAAAAAAAAAAAAAAAAAAAA 196

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
              |||   |  | || || | | |     ||| ||||  | || |  |  |||   |
Db        197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 256

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
              | || || | | |       ||| ||||    || || || |||   |   | || ||
Db        257 AAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAA 316

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
              | | |      |||  |||     || || || |||   |   | || || | | |
Db        317 AAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
              |||  ||||     || || || |||   |   | || || | | |      ||| |||
Db        377 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 436

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
              |   | || |  |  |||   |   | || || | | |      ||| ||||     ||
Db        437 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 496

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
              || || |||   |   | || || | | |     ||| ||||     || || || |||
Db        497 AAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 556

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
              |   | || || | | |      ||| ||||     || || || |||   |   |
Db        557 AAAANAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAANAAAAA 616

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
              || || | |  |||  ||||     || || || |||   |   | || || | |
Db        617 AAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 676

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
              |      ||| ||||    || || || |     |   | || || | |      |
Db        677 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNAAAANAAAAAAAAAAAAAAAAAAAA 736

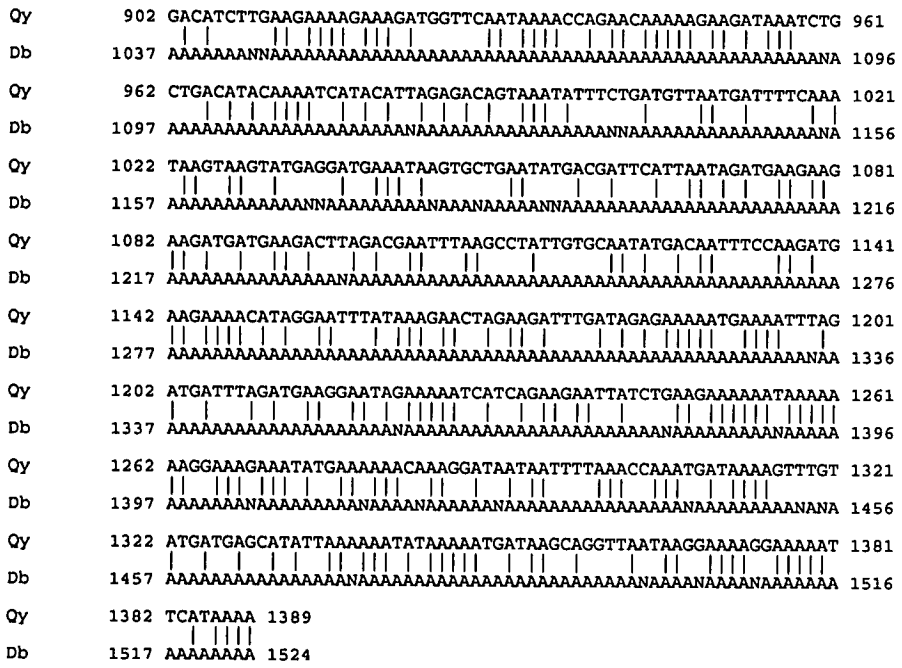
Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
              || ||||     || |  |  ||| | |   | || |||  ||      | |  |||||||
Db        737 AAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAA 796

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
              |   | ||| ||||||  || |     | | |  | | |  | |  |      |
Db        797 AAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 856

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
              ||| |  |   | | ||     |  | ||||  ||      | | | | |  | ||
Db        857 AAAAAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAA 916

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
                  |    | | || | |   || |  | | |     ||  |||     | || |
Db        917 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 976

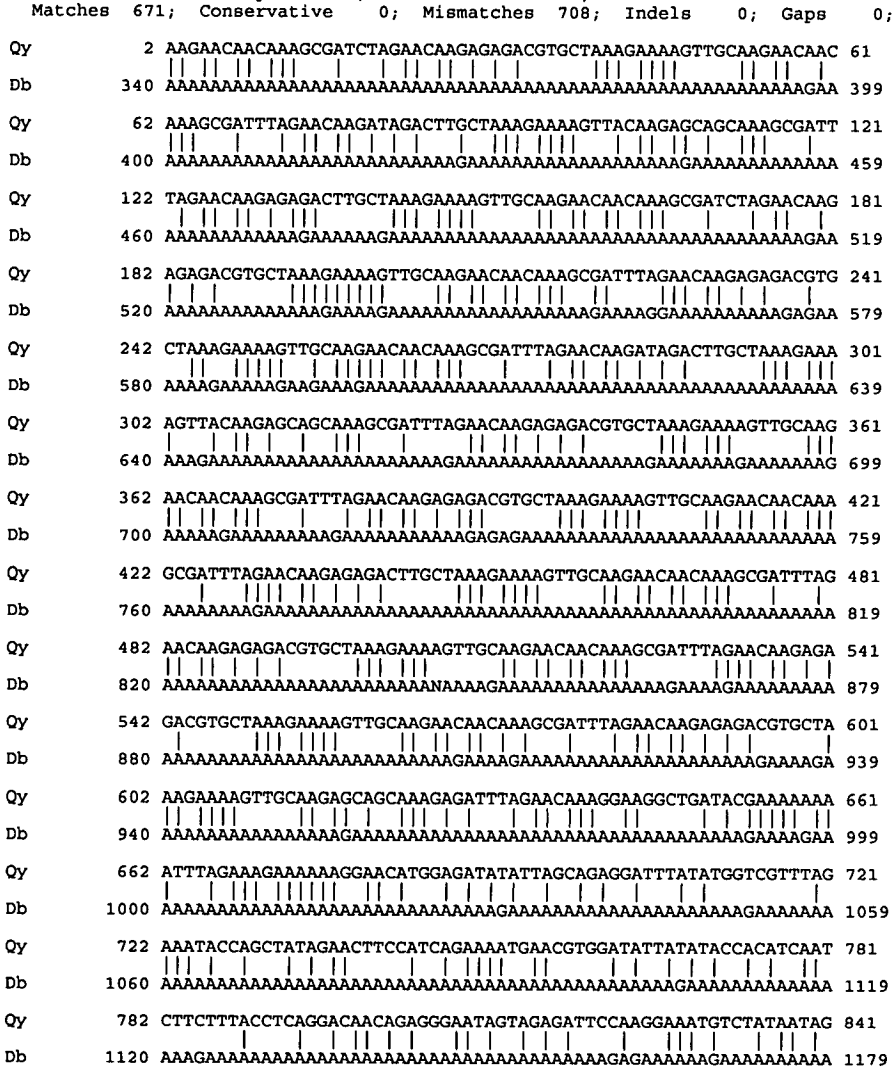
Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
              |||||| ||| | | ||   |  | || |||     ||  |  ||   | | | | |||
Db        977 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 1036
```



RESULT 3
DR125630
LOCUS DR125630 1808 bp mRNA linear EST 15-JUN-2005
DEFINITION 49096520 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O23 5', mRNA sequence.
ACCESSION DR125630
VERSION DR125630.1 GI:67842328
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226697953

Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1808
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O23"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

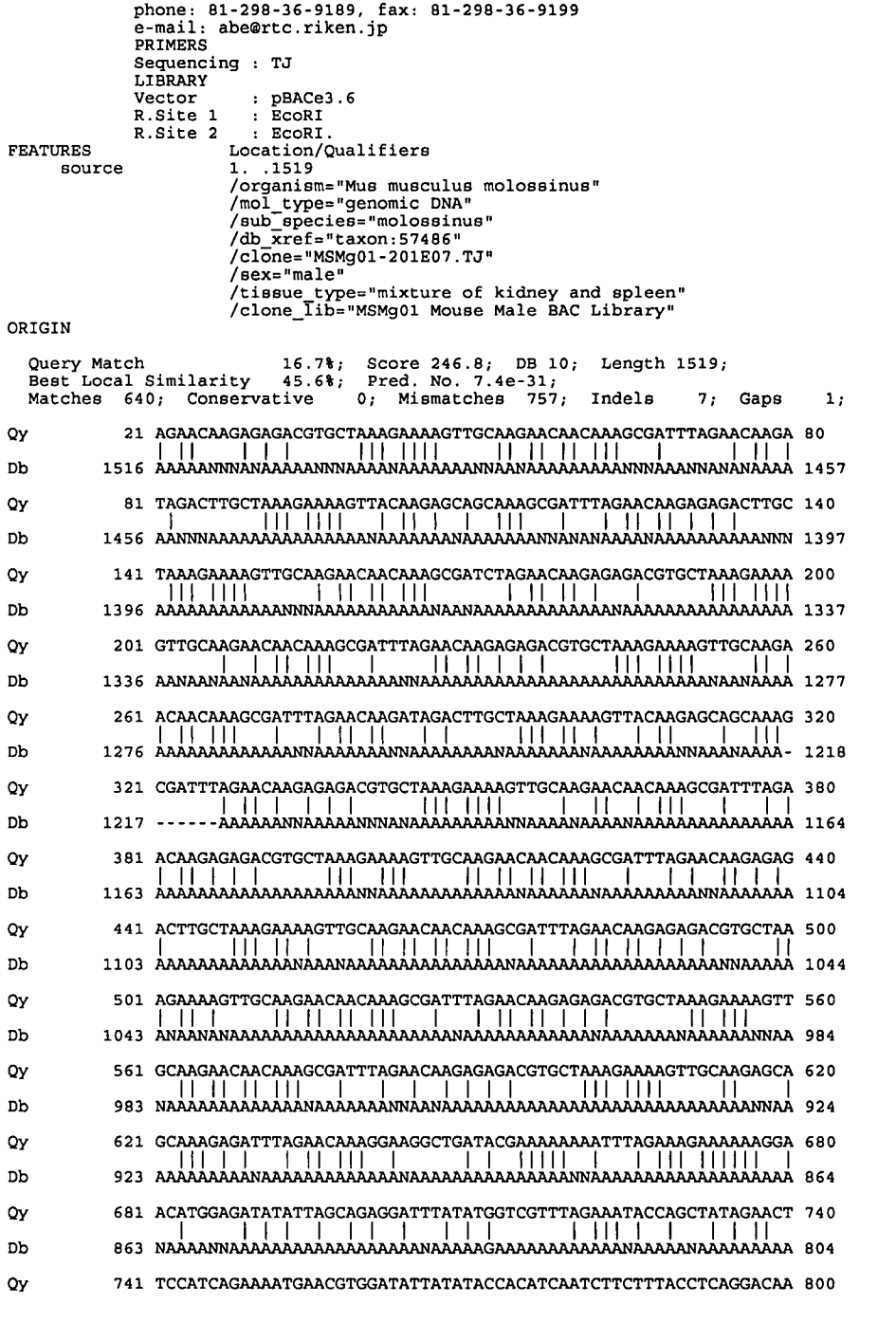
Query Match 16.9%; Score 251; DB 8; Length 1808;
Best Local Similarity 48.7%; Pred. No. 1.5e-31;
Matches 671; Conservative 0; Mismatches 708; Indels 0; Gaps 0;

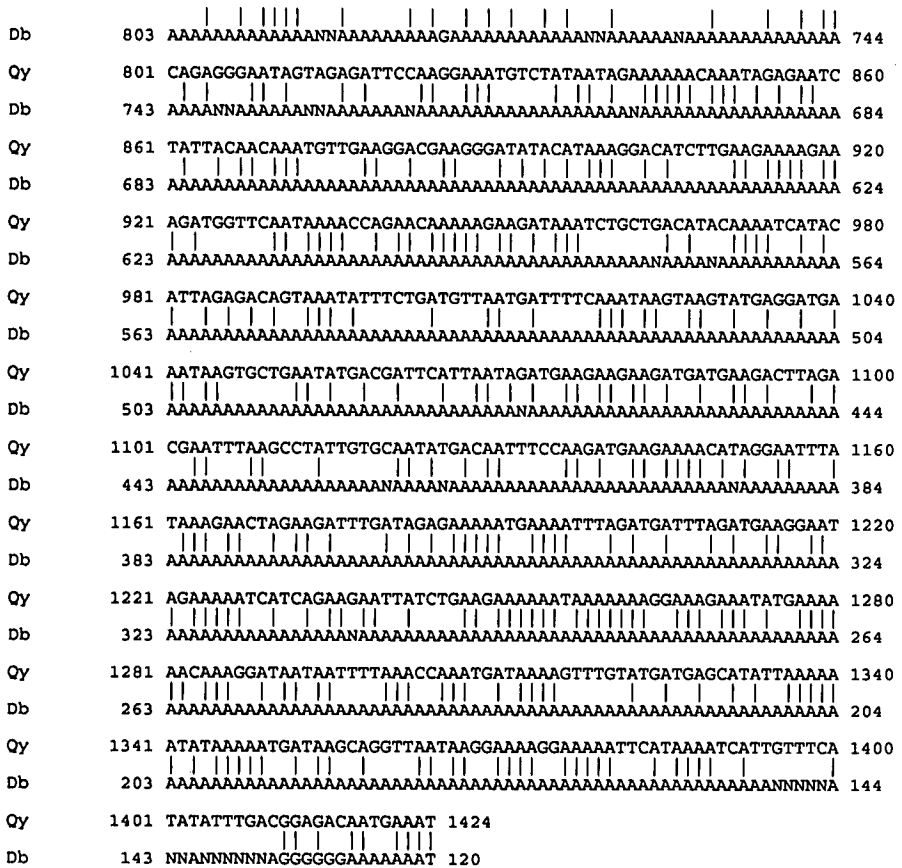




```
RESULT 4
AG386893/c
LOCUS       AG386893               1519 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-201E07.TJ, genomic survey
            sequence.
ACCESSION   AG386893
VERSION     AG386893.1  GI:47998098
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1519)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1519
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-201E07.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

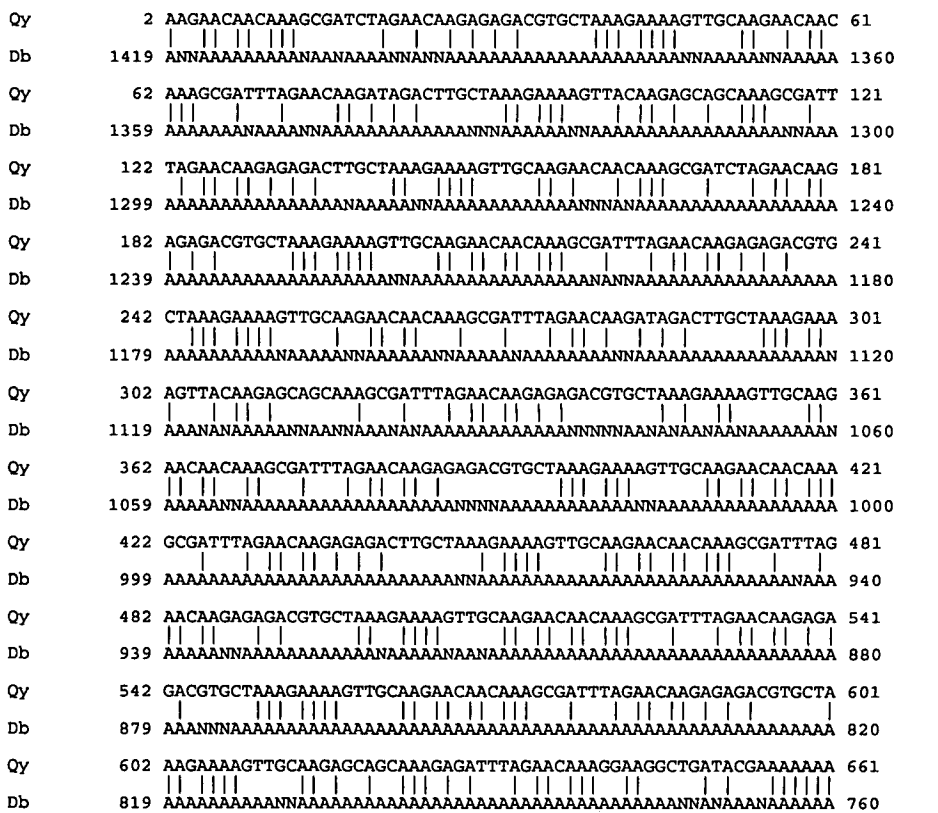
Query Match 16.7%; Score 246.8; DB 10; Length 1519;
Best Local Similarity 45.6%; Pred. No. 7.4e-31;
Matches 640; Conservative 0; Mismatches 757; Indels 7; Gaps 1;

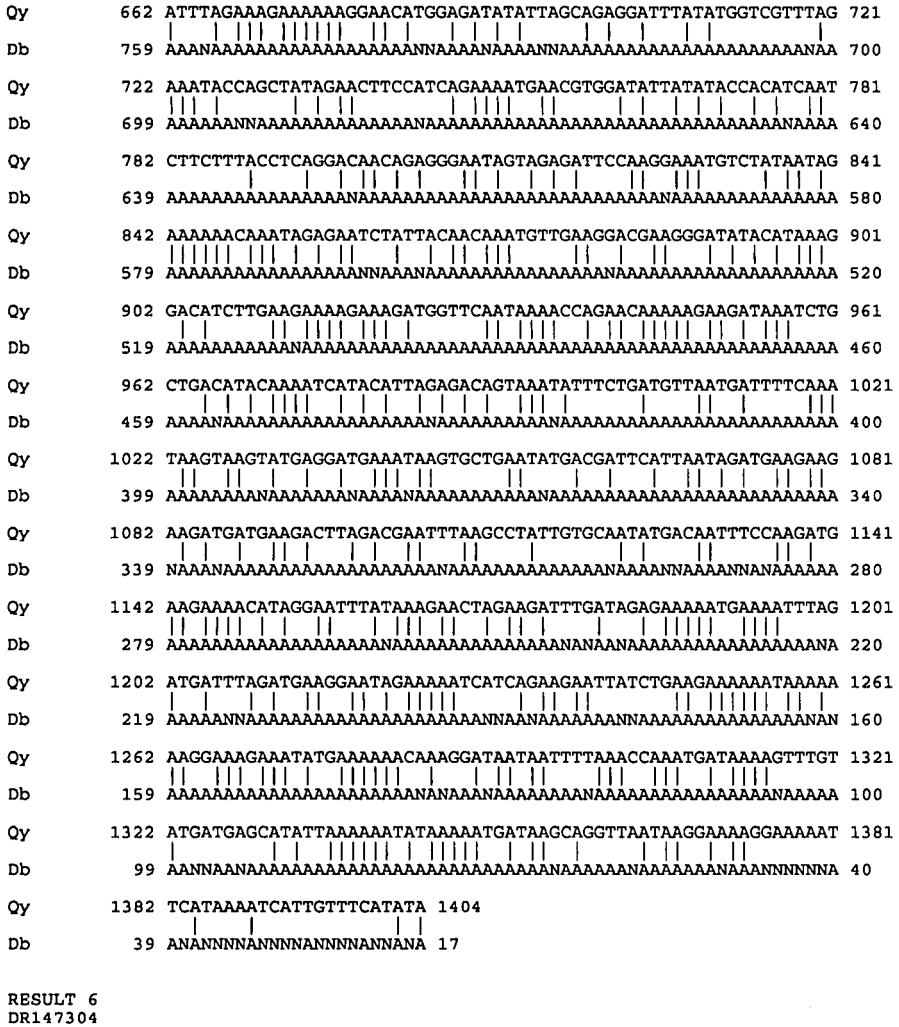




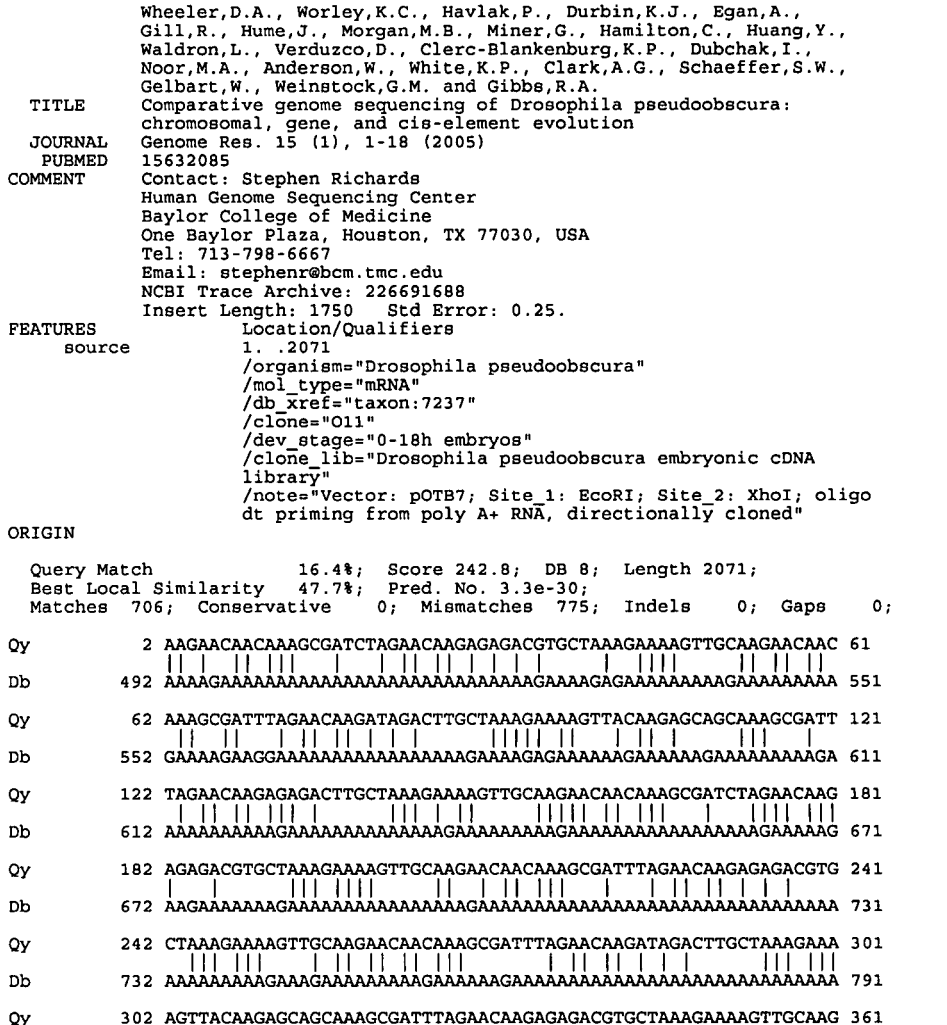
RESULT 5
AJ591978/c
LOCUS AJ591978 1453 bp DNA linear GSS 15-JAN-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 598D01, genomic survey sequence.
ACCESSION AJ591978
VERSION AJ591978.1 GI:37941602
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1453)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).
FEATURES Location/Qualifiers
source 1. .1453
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="598D01"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Wassilewskija"
misc_feature 1. .1453
/note="T-DNA flanking sequence right border"
ORIGIN

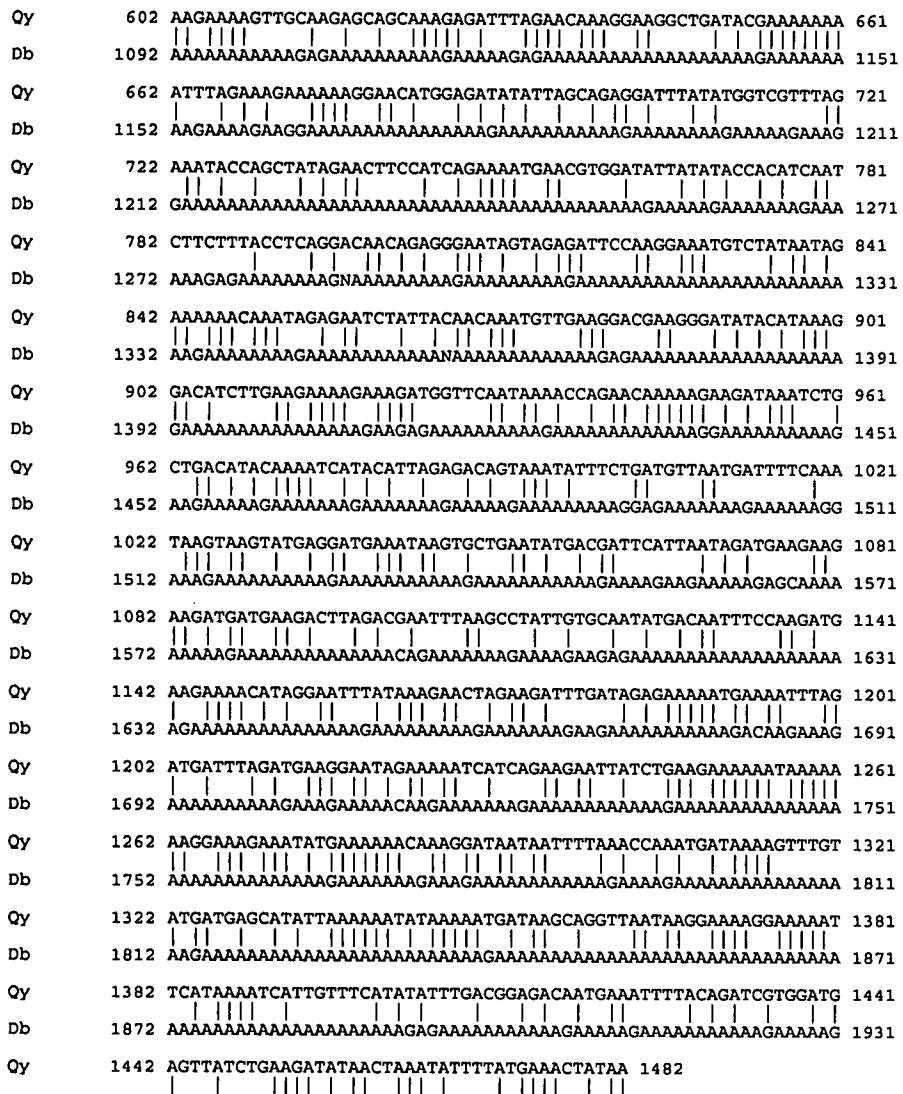
Query Match 16.6%; Score 245.6; DB 10; Length 1453;
Best Local Similarity 44.0%; Pred. No. 1.2e-30;
Matches 617; Conservative 0; Mismatches 786; Indels 0; Gaps 0;





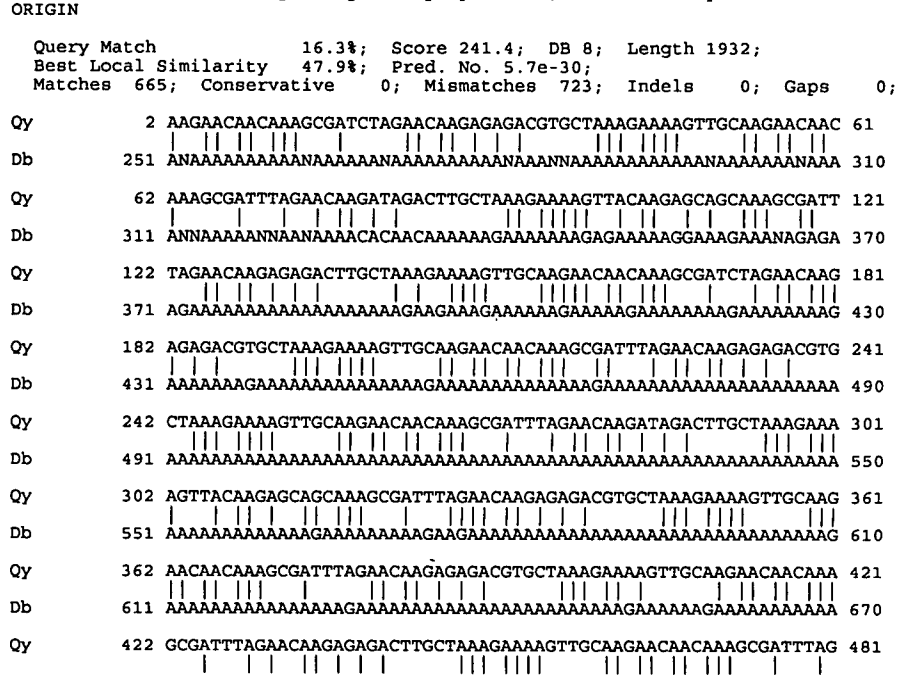
RESULT 6
DR147304
LOCUS DR147304 2071 bp mRNA linear EST 16-JUN-2005
DEFINITION 49027161 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O11 3', mRNA sequence.
ACCESSION DR147304
VERSION DR147304.1 GI:67892394
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2071)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226691688
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .2071
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O11"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

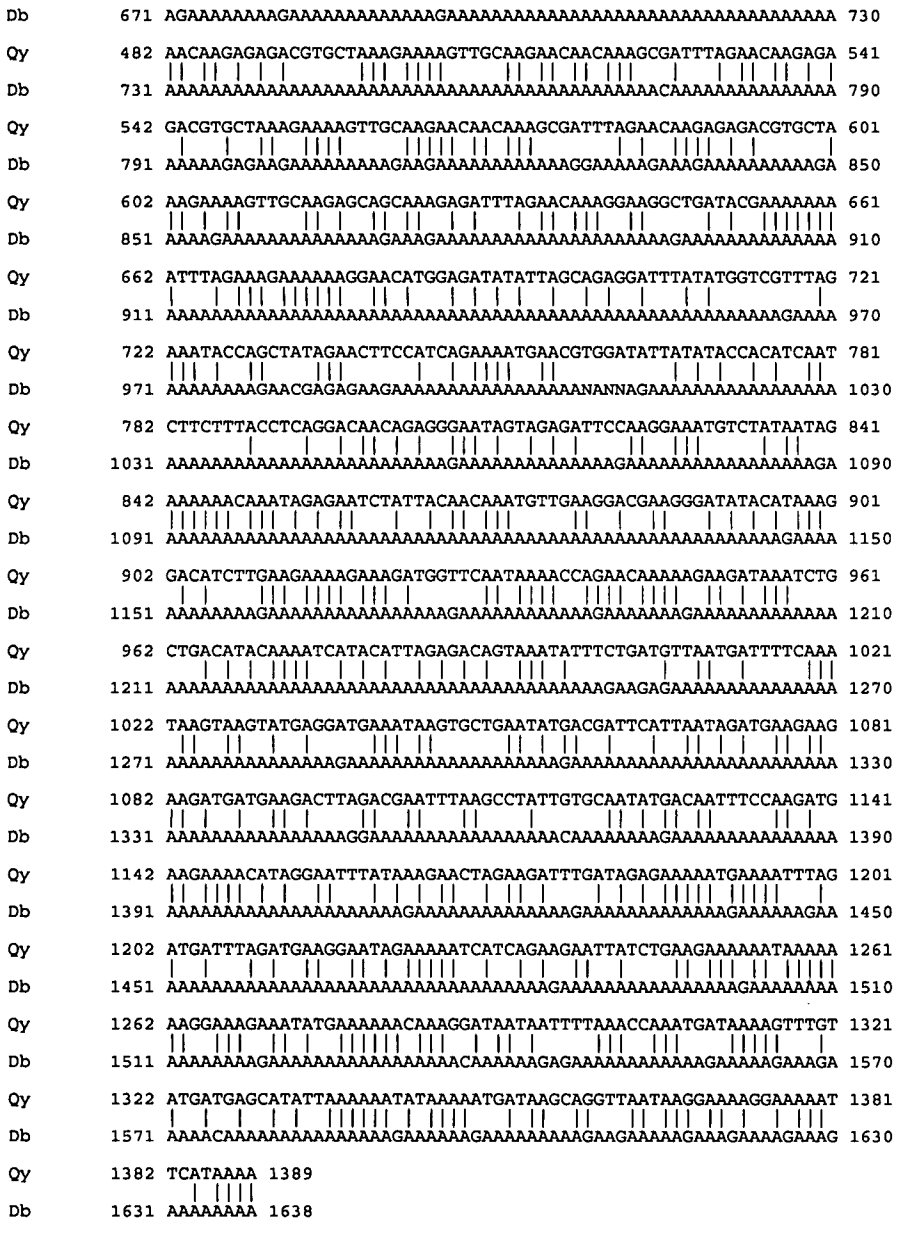




RESULT 7
DR141720
LOCUS DR141720 1932 bp mRNA linear EST 16-JUN-2005
DEFINITION 49300102 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone I13 3', mRNA sequence.
ACCESSION DR141720
VERSION DR141720.1 GI:67886810
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1932)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226717045
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1932
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="I13"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

Query Match 16.3%; Score 241.4; DB 8; Length 1932;
Best Local Similarity 47.9%; Pred. No. 5.7e-30;
Matches 665; Conservative 0; Mismatches 723; Indels 0; Gaps 0;

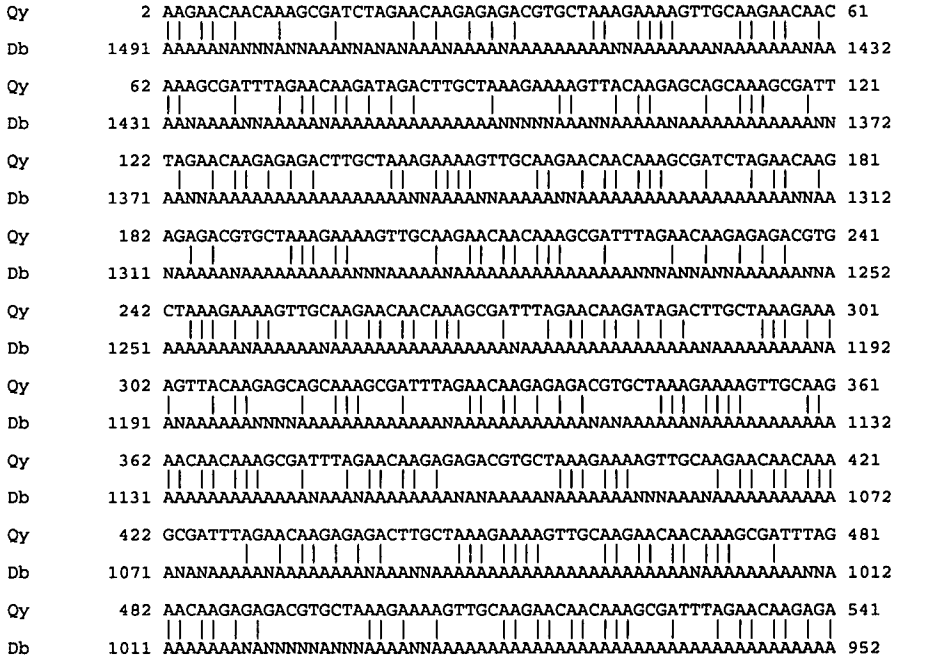


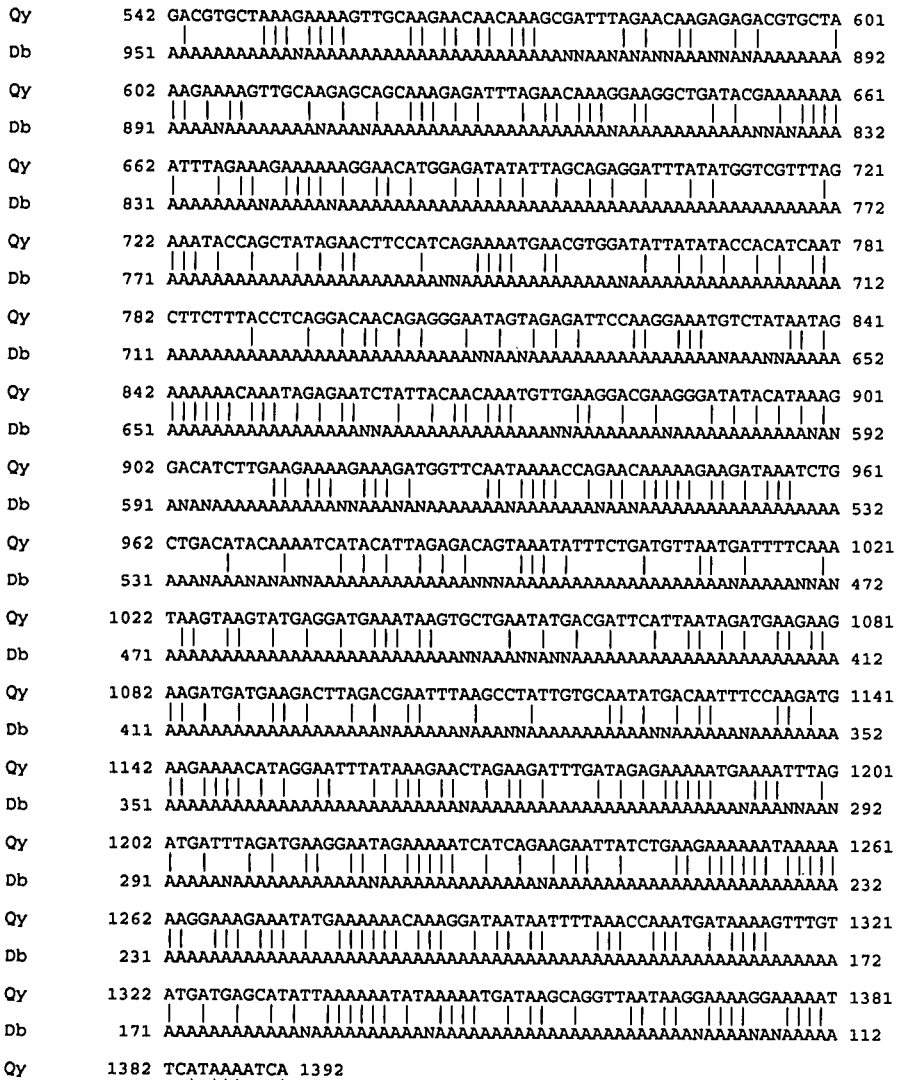


RESULT 8
CG753221/c
LOCUS CG753221 1491 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-2-A01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753221

VERSION CG753221.1 GI:37977480
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1491
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

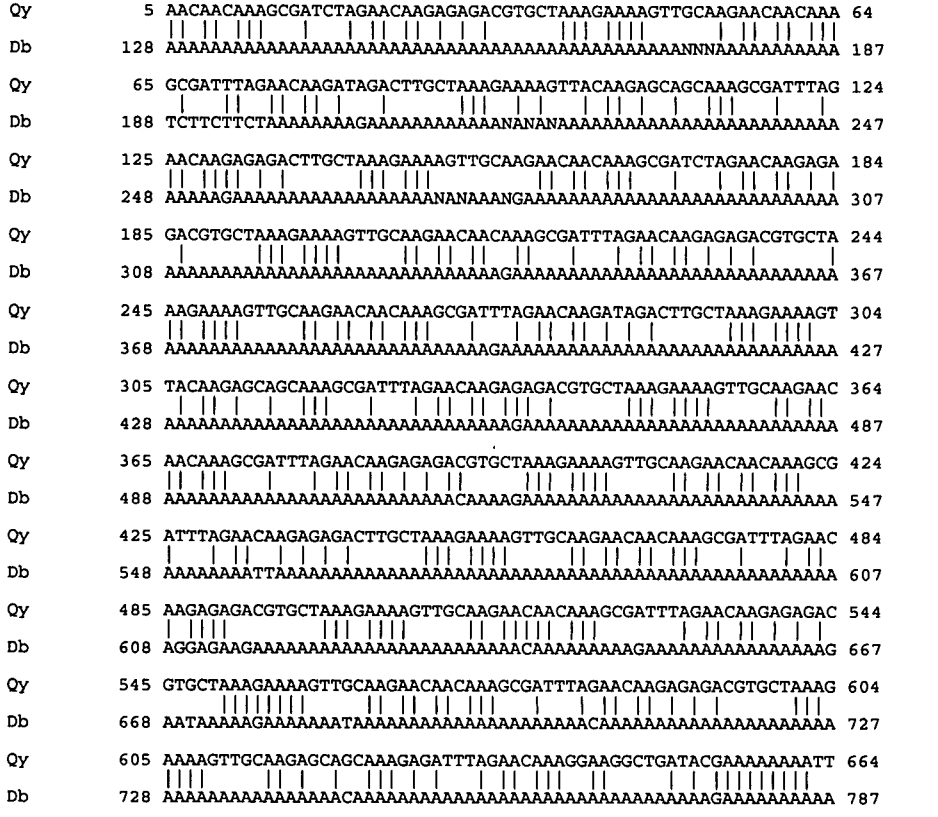
Query Match 16.2%; Score 240; DB 10; Length 1491;
Best Local Similarity 43.4%; Pred. No. 9.8e-30;
Matches 603; Conservative 0; Mismatches 788; Indels 0; Gaps 0;

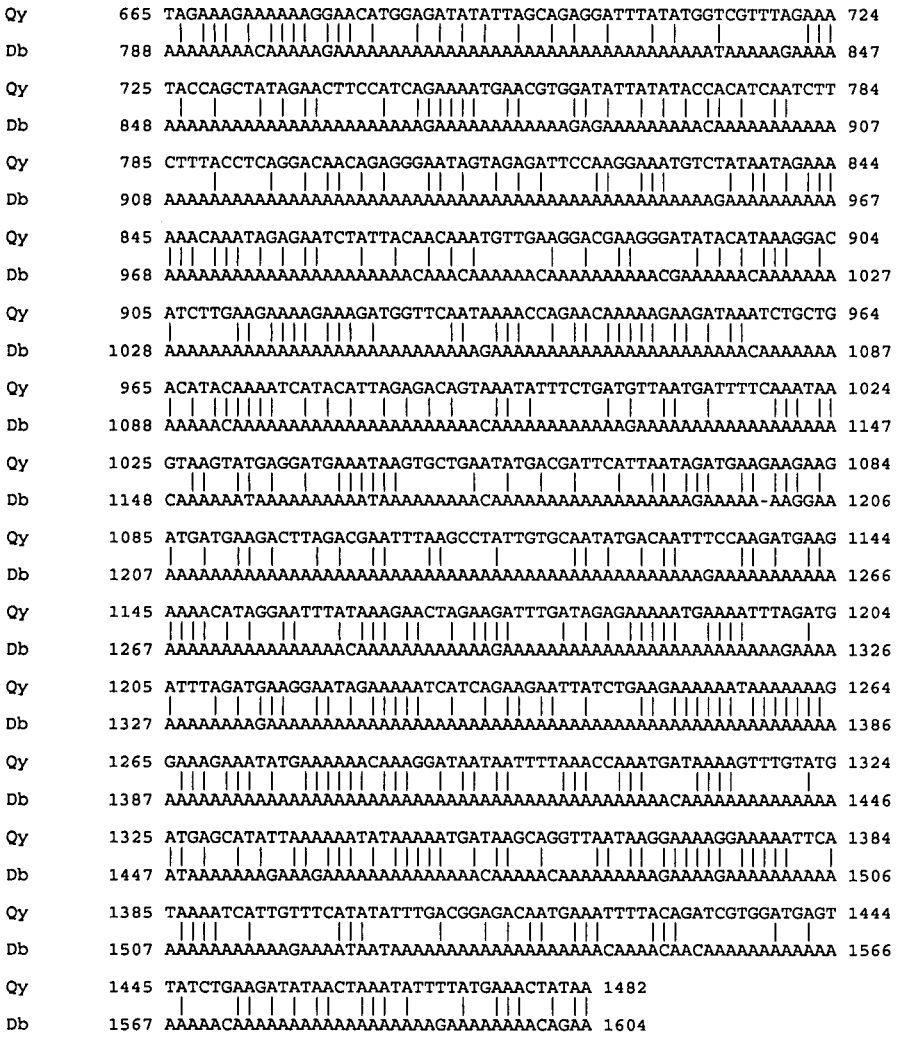




RESULT 9
CL114085
LOCUS CL114085 1727 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-59J11_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59J11, genomic survey sequence.
ACCESSION CL114085
VERSION CL114085.1 GI:40607720
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1727)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 280
High quality sequence stop: 1423.
FEATURES Location/Qualifiers
source 1. .1727
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-59J11"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

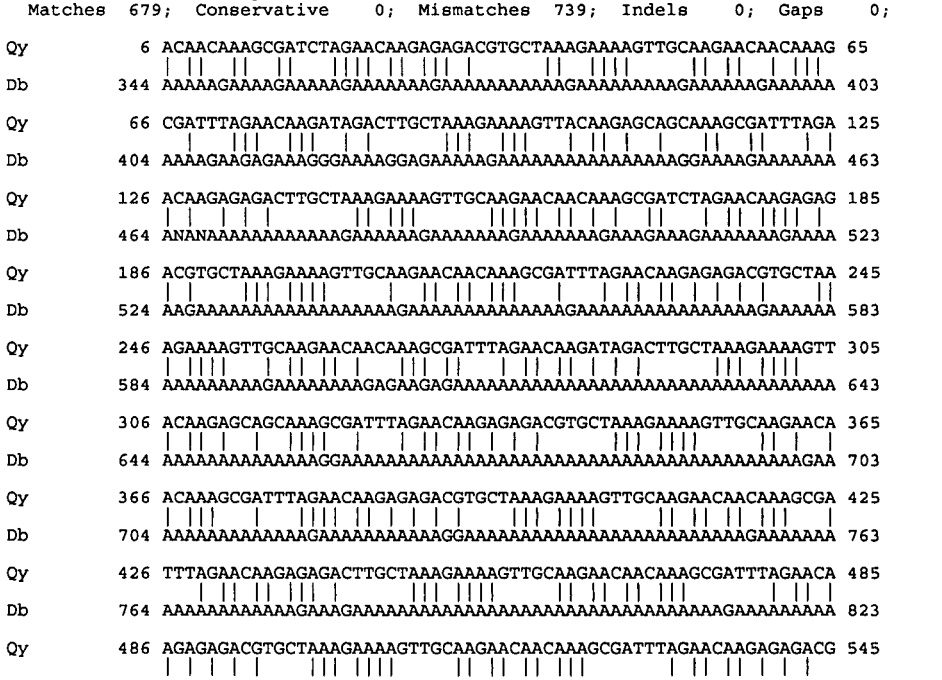
Query Match 16.2%; Score 239.4; DB 10; Length 1727;
Best Local Similarity 47.8%; Pred. No. 1.2e-29;
Matches 707; Conservative 0; Mismatches 770; Indels 1; Gaps 1;

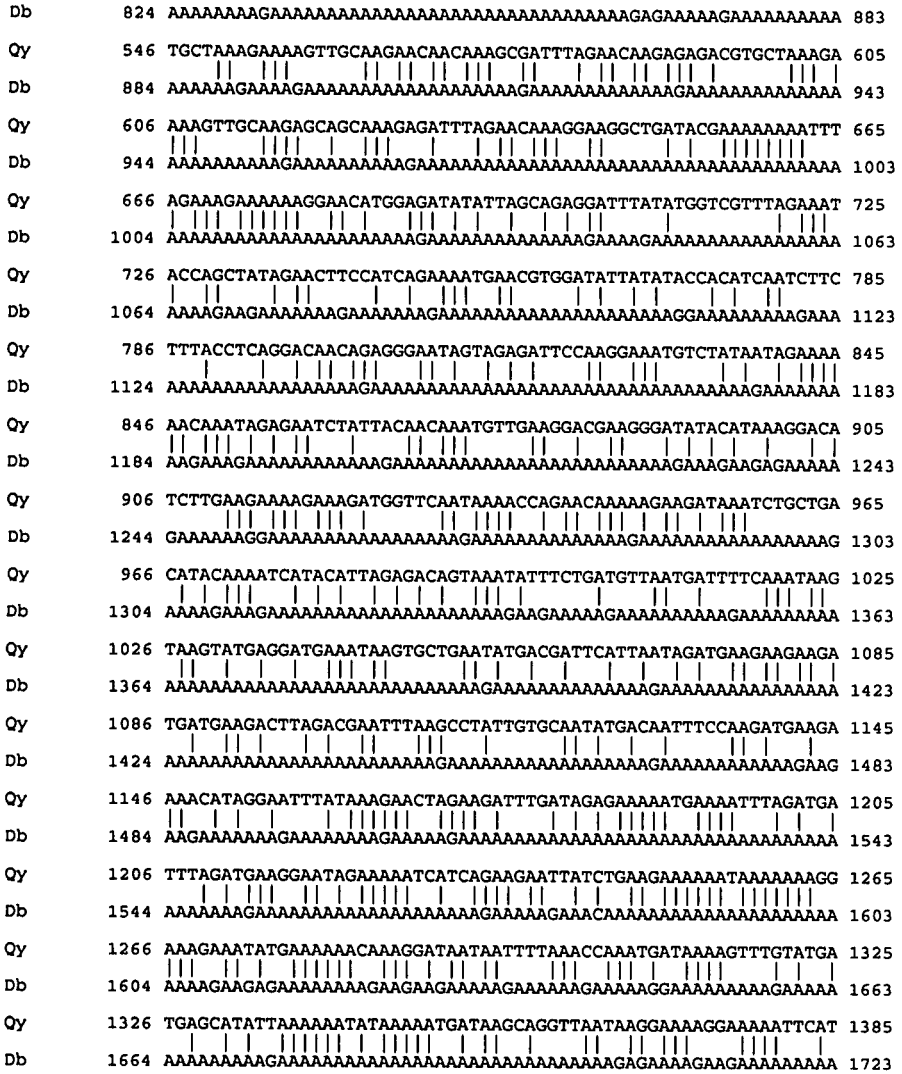




RESULT 10
DR141924
LOCUS DR141924 1783 bp mRNA linear EST 16-JUN-2005
DEFINITION 49190649 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone O7 3', mRNA sequence.
ACCESSION DR141924
VERSION DR141924.1 GI:67887014
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1783)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226707030
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1783
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="O7"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

Query Match 16.0%; Score 236.8; DB 8; Length 1783;
Best Local Similarity 47.9%; Pred. No. 3.3e-29;
Matches 679; Conservative 0; Mismatches 739; Indels 0; Gaps 0;

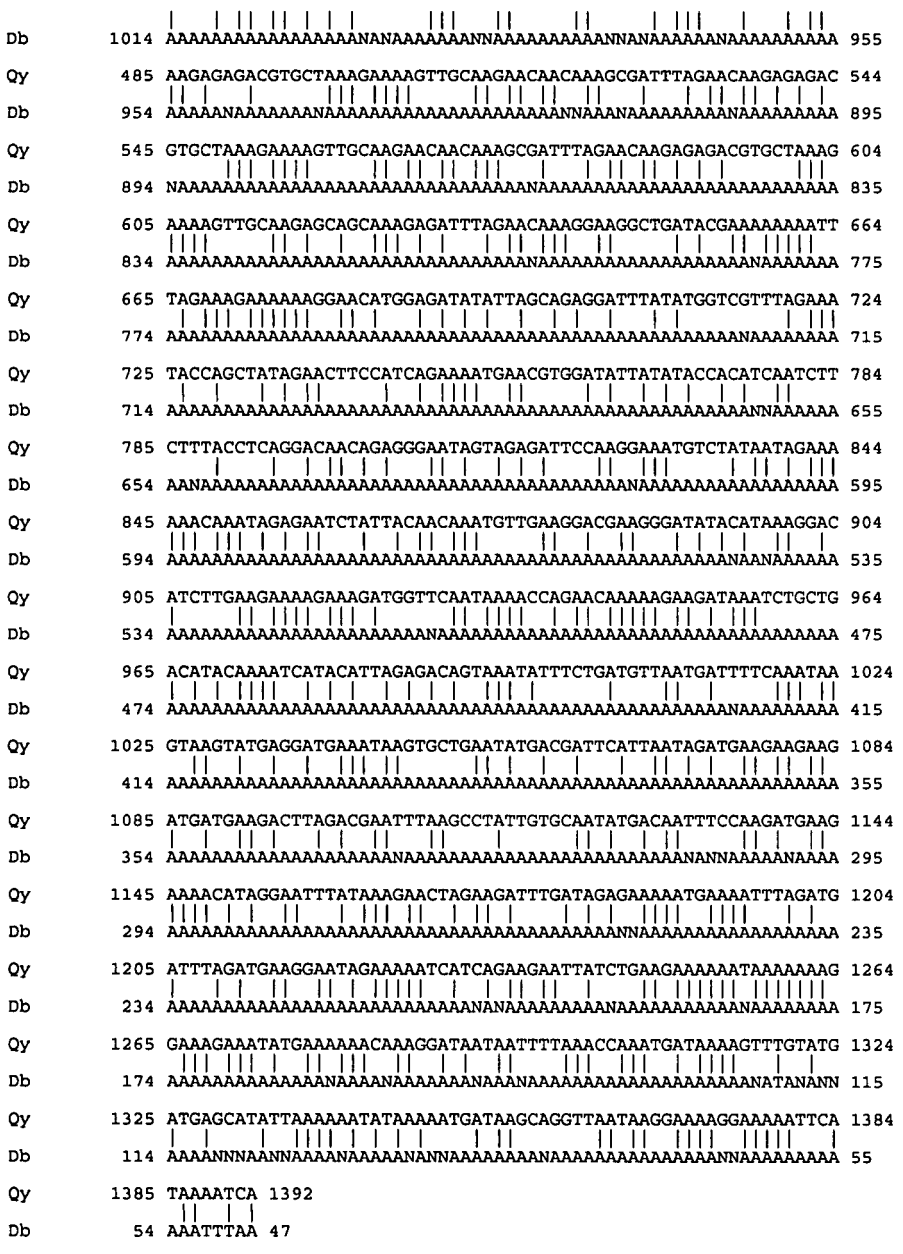




RESULT 11
AJ592058/c
LOCUS AJ592058 1434 bp DNA linear GSS 15-JAN-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 602D05, genomic survey sequence.
ACCESSION AJ592058
VERSION AJ592058.1 GI:37941682
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1434)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).
FEATURES Location/Qualifiers
source 1. .1434
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="602D05"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Wassilewskija"
misc_feature 1. .1434
/note="T-DNA flanking sequence right border"
ORIGIN

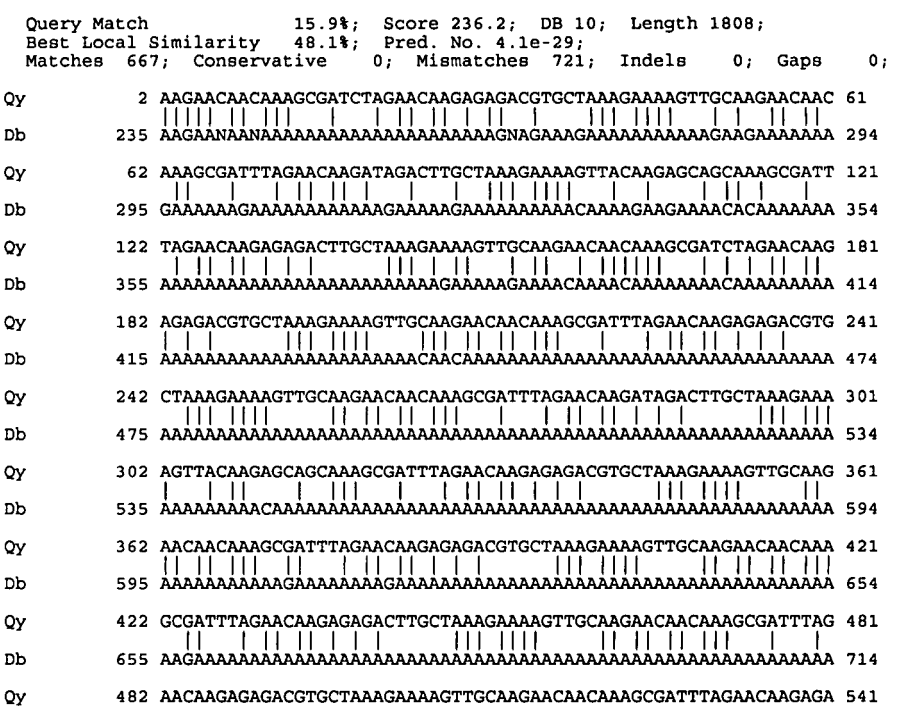
Query Match 15.9%; Score 236.2; DB 10; Length 1434;
Best Local Similarity 44.4%; Pred. No. 4.1e-29;
Matches 616; Conservative 0; Mismatches 772; Indels 0; Gaps 0;

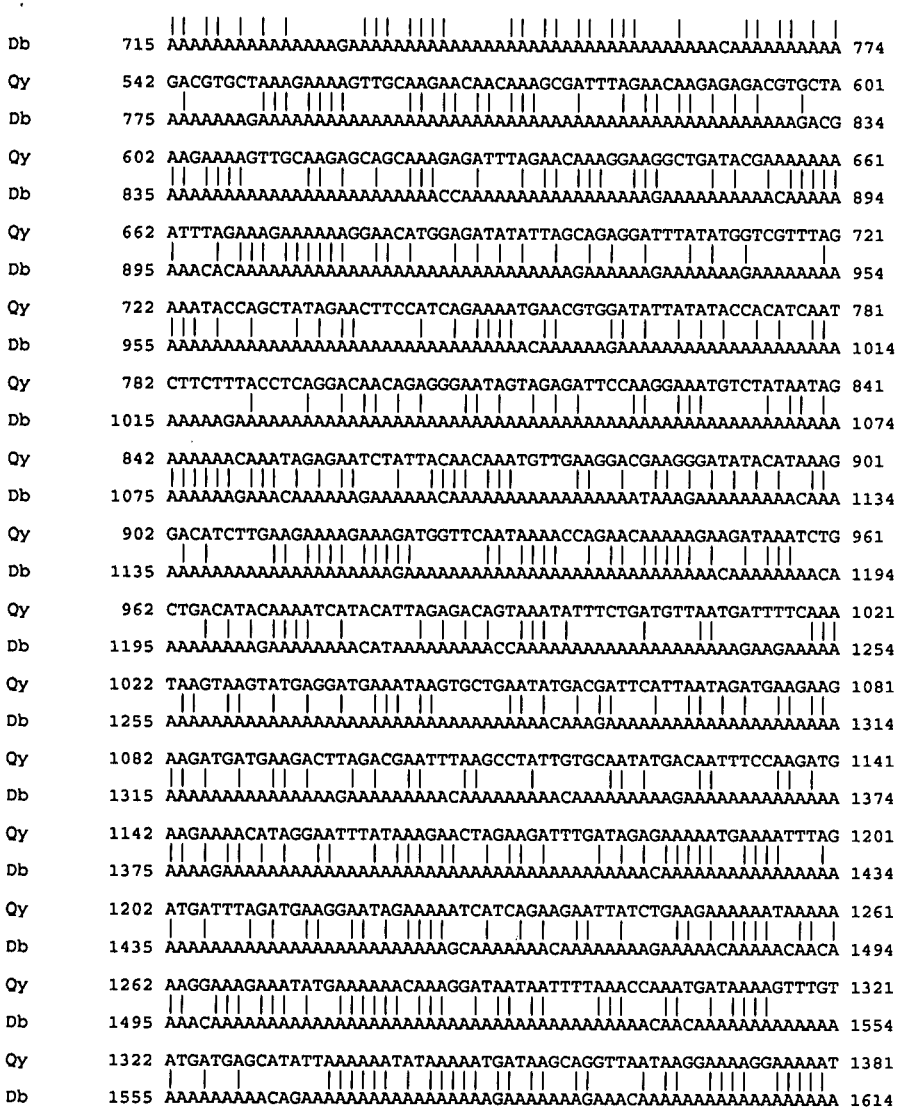




RESULT 12
CL078963
LOCUS CL078963 1808 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-154E11_RM4.1 CH216 Xenopus tropicalis genomic clone
CH216-154E11, genomic survey sequence.

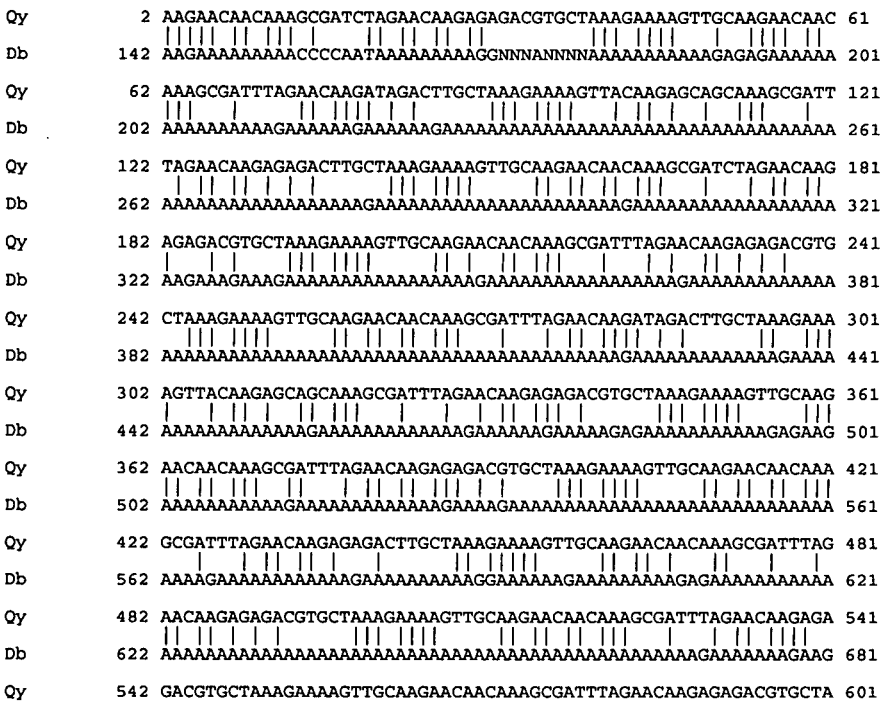
ACCESSION CL078963
VERSION CL078963.1 GI:40534876
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggcatcggtcgagc
Class: BAC ends
High quality sequence start: 408
High quality sequence stop: 936.
FEATURES Location/Qualifiers
source 1. .1808
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-154E11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

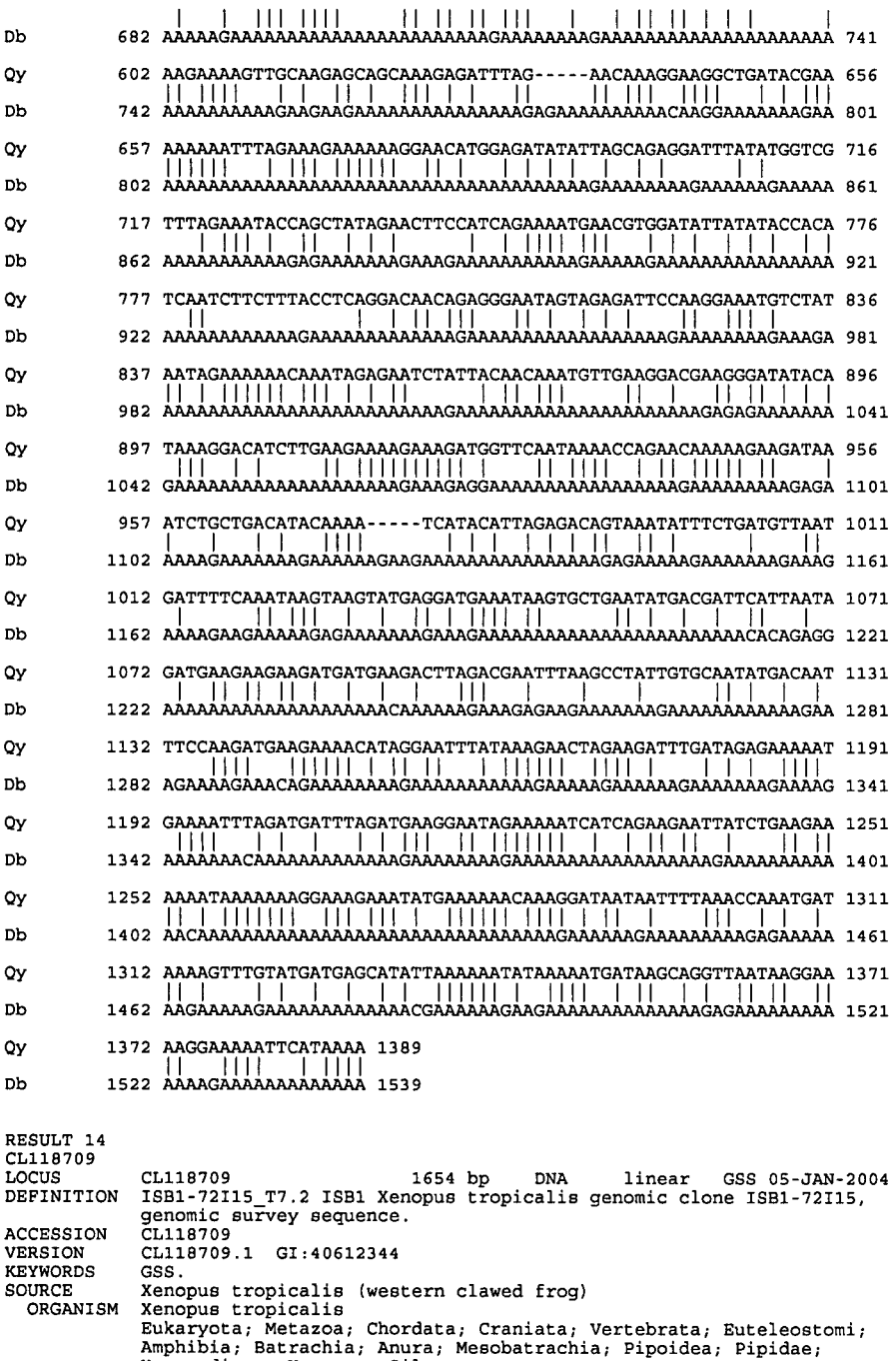




RESULT 13
CL066150
LOCUS CL066150 1797 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-108B7_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-108B7, genomic survey sequence.
ACCESSION CL066150
VERSION CL066150.1 GI:40522063
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 306
High quality sequence stop: 453.
FEATURES Location/Qualifiers
source 1. .1797
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-108B7"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 15.9%; Score 235.8; DB 10; Length 1797;
Best Local Similarity 48.9%; Pred. No. 4.8e-29;
Matches 684; Conservative 0; Mismatches 704; Indels 10; Gaps 2;

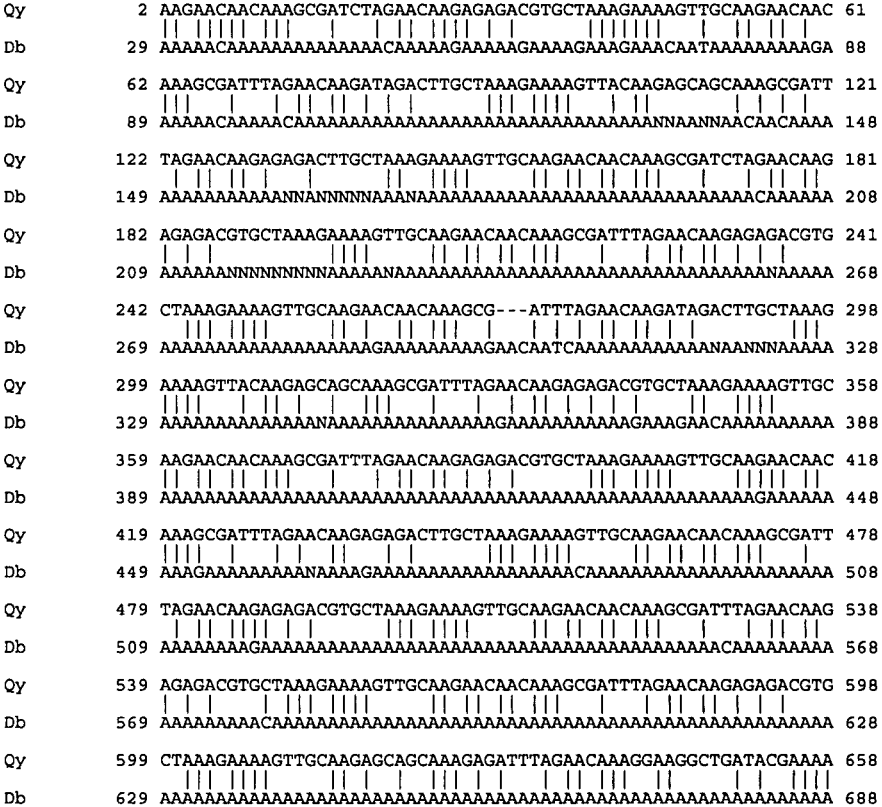


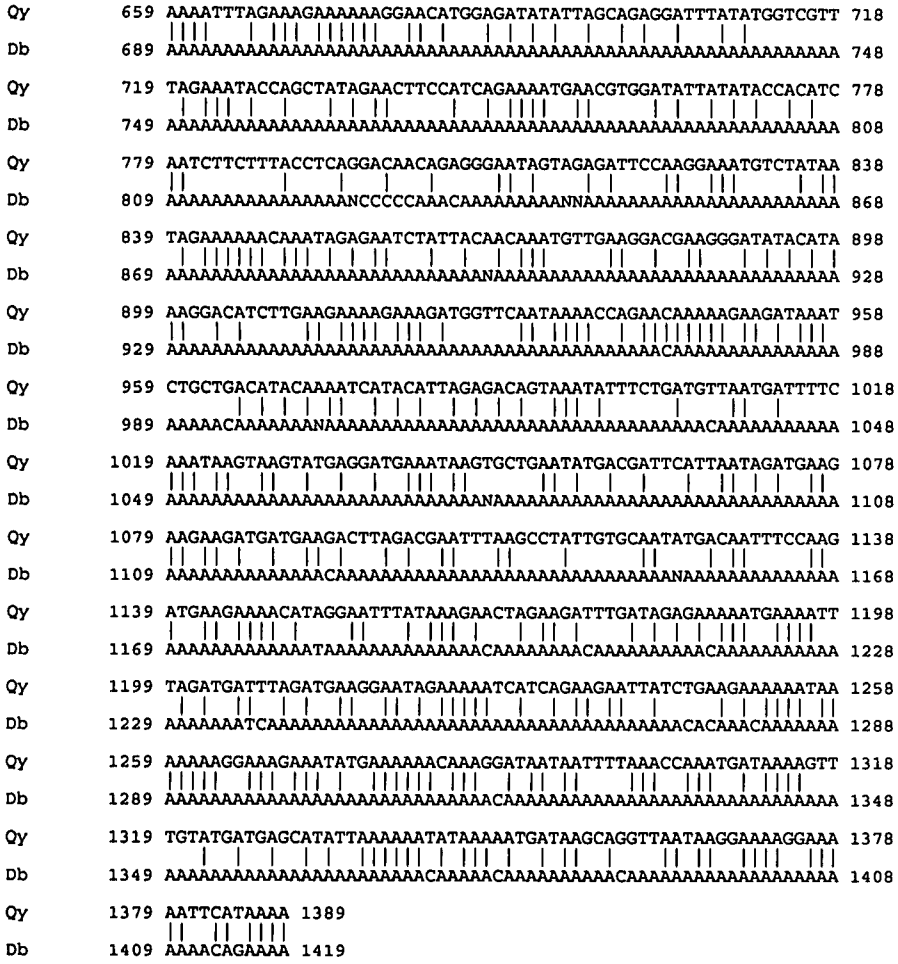


RESULT 14
CL118709
LOCUS CL118709 1654 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-72I15_T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72I15, genomic survey sequence.
ACCESSION CL118709
VERSION CL118709.1 GI:40612344
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 1654)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 468
High quality sequence stop: 812.
FEATURES Location/Qualifiers
source 1. .1654
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-72I15"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 15.7%; Score 233.4; DB 10; Length 1654;
Best Local Similarity 47.5%; Pred. No. 1.2e-28;
Matches 661; Conservative 0; Mismatches 727; Indels 3; Gaps 1;





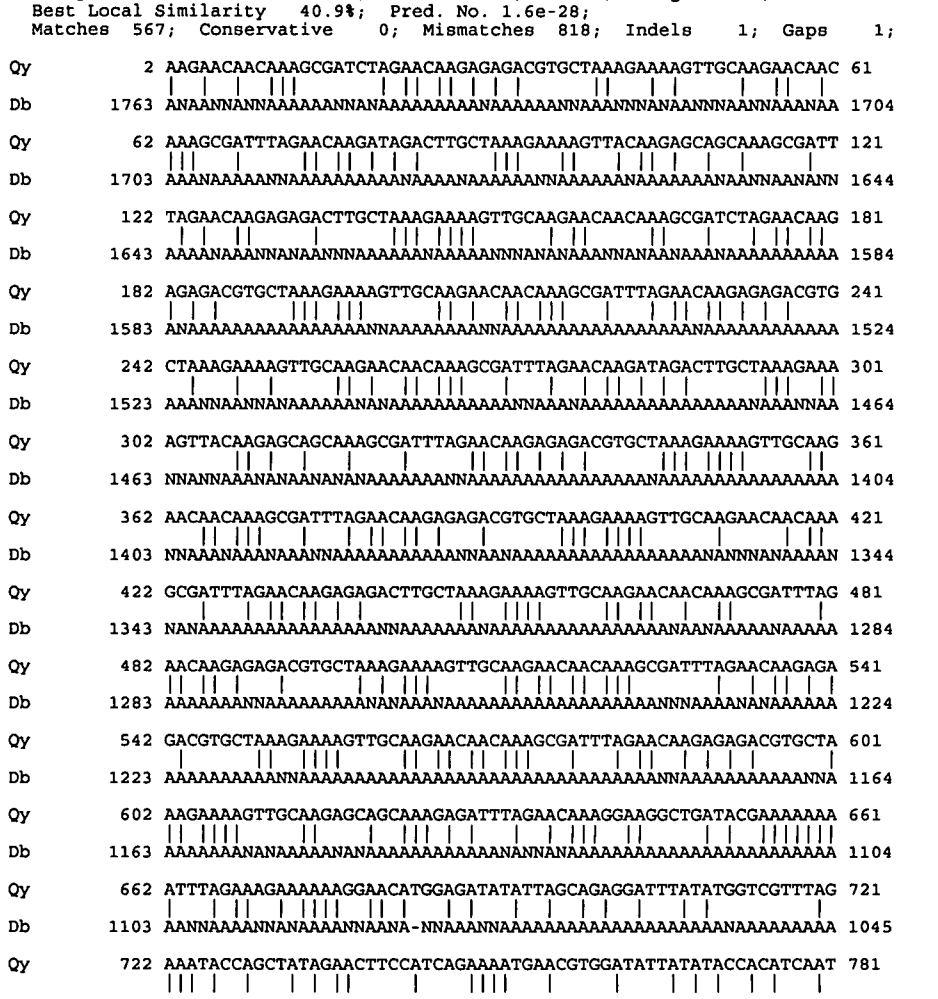
RESULT 15
CG753732/c
LOCUS CG753732 1811 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-4-G03.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753732
VERSION CG753732.1 GI:37978509
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1811)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ

Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .1811
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

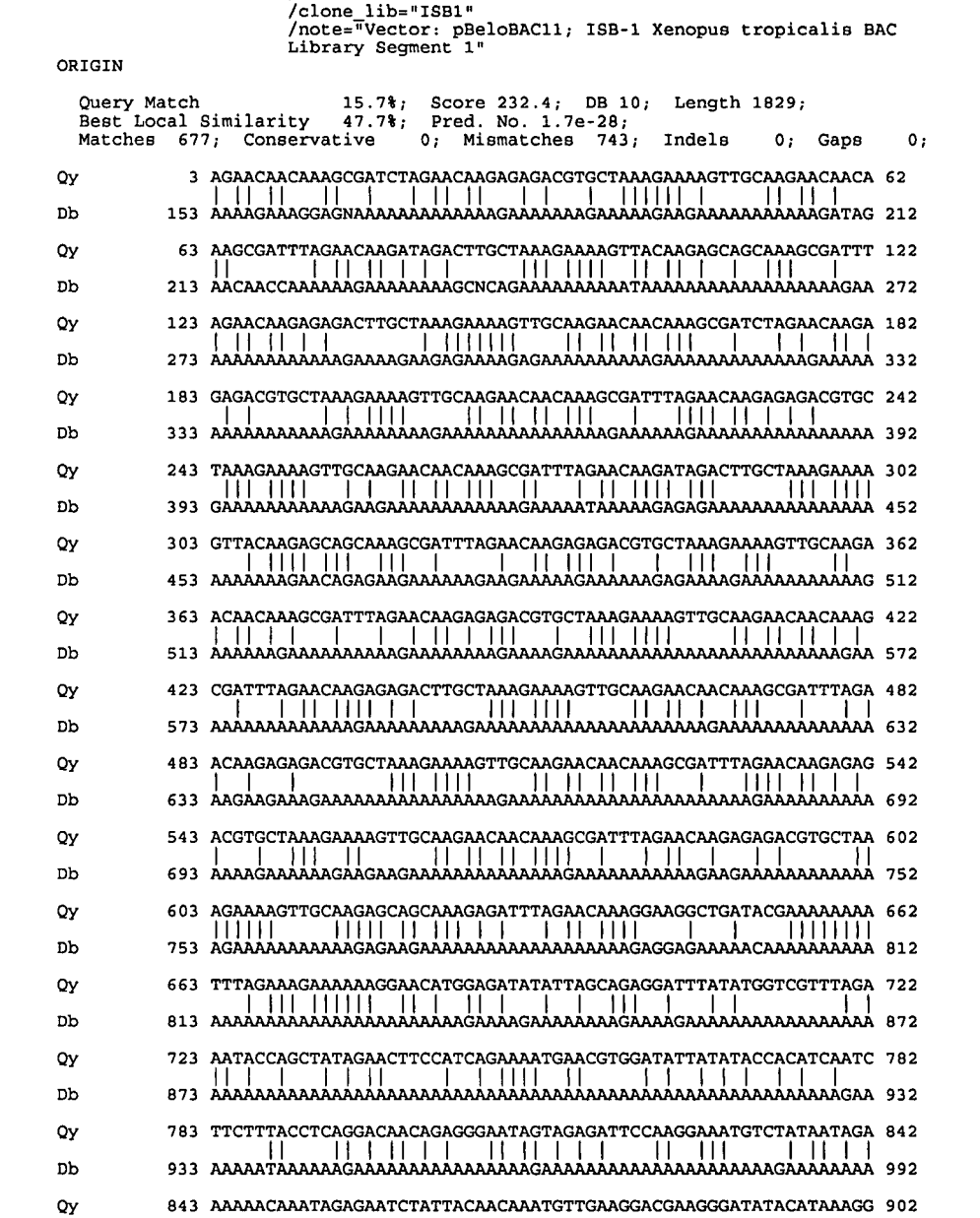
ORIGIN

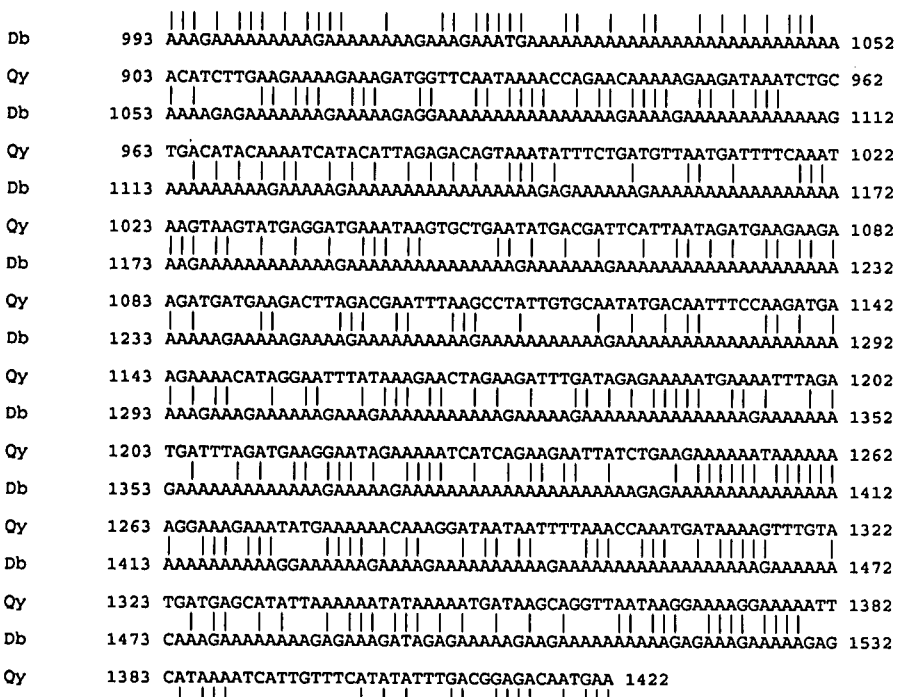
Query Match 15.7%; Score 232.6; DB 10; Length 1811;
Best Local Similarity 40.9%; Pred. No. 1.6e-28;
Matches 567; Conservative 0; Mismatches 818; Indels 1; Gaps 1;





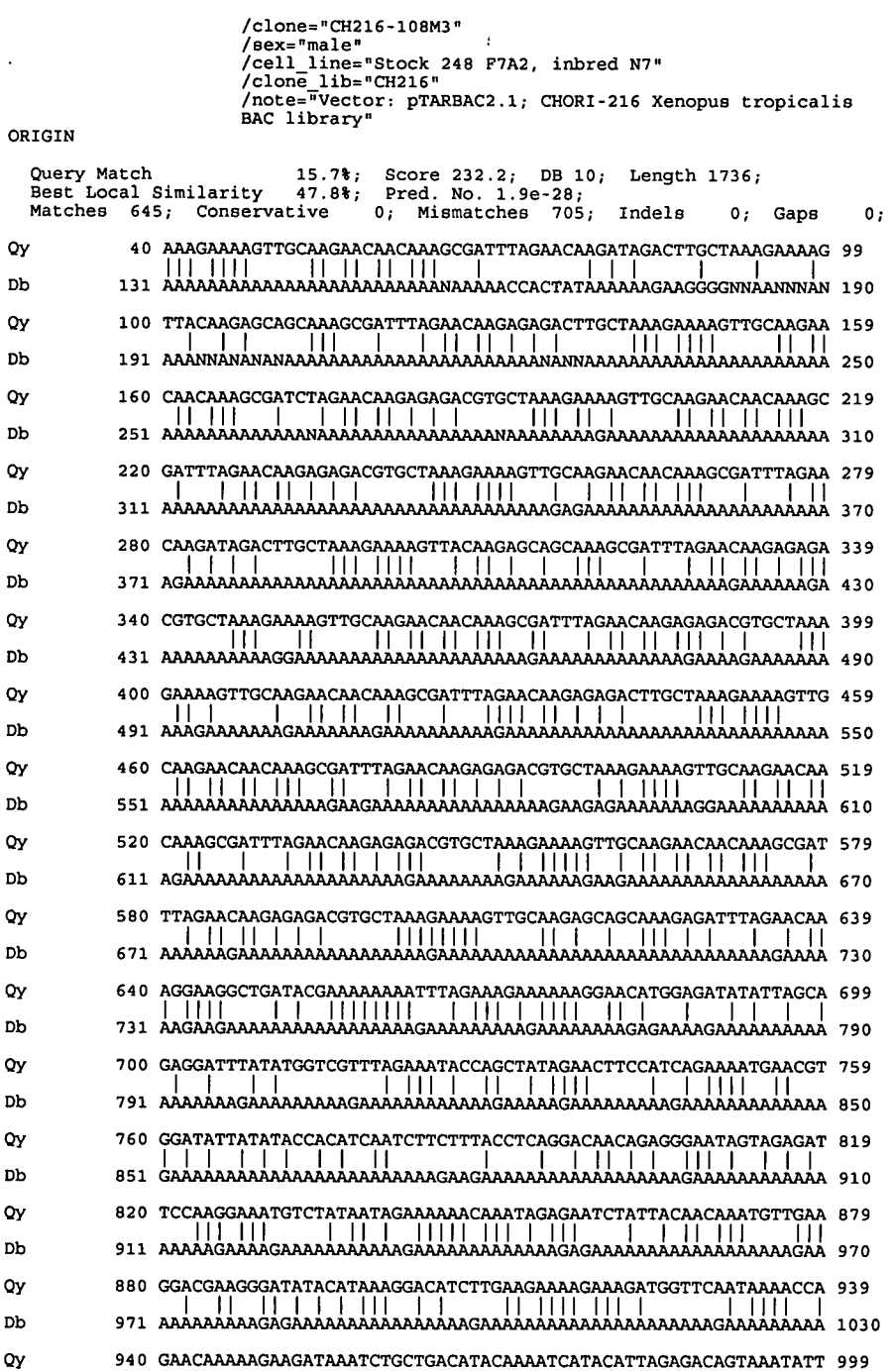
```
RESULT 16
CL109409
LOCUS       CL109409               1829 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-51H21_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-51H21,
            genomic survey sequence.
ACCESSION   CL109409
VERSION     CL109409.1  GI:40603044
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1829)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 856
            High quality sequence stop: 936.
FEATURES             Location/Qualifiers
     source          1. .1829
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-51H21"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN
```

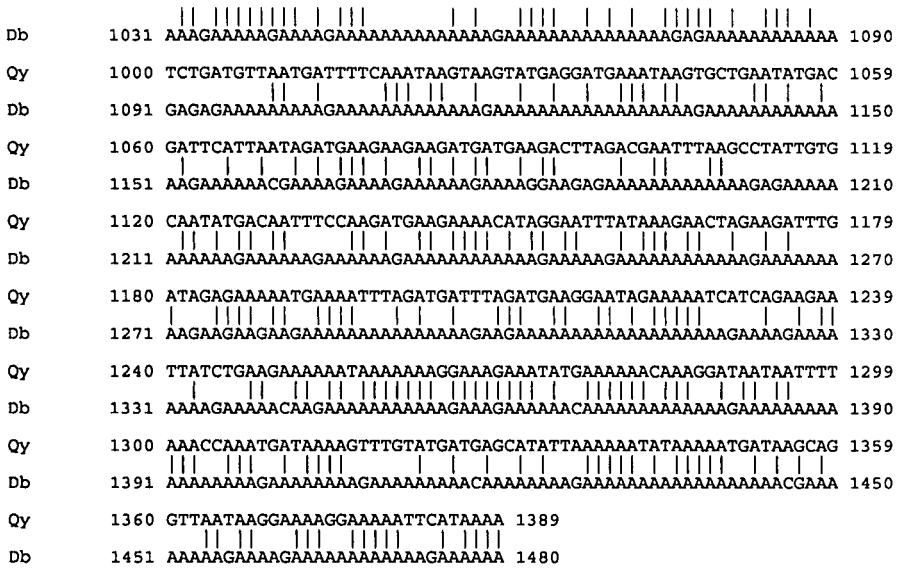




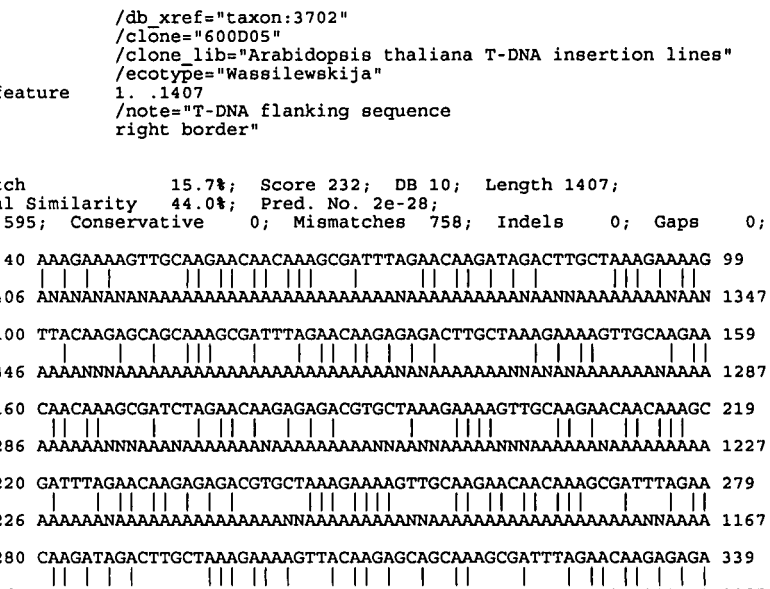
```
RESULT 17
CL066373
LOCUS       CL066373               1736 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-108M3_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-108M3, genomic survey sequence.
ACCESSION   CL066373
VERSION     CL066373.1  GI:40522286
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1736)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 525
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1. .1736
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-108M3"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN
```

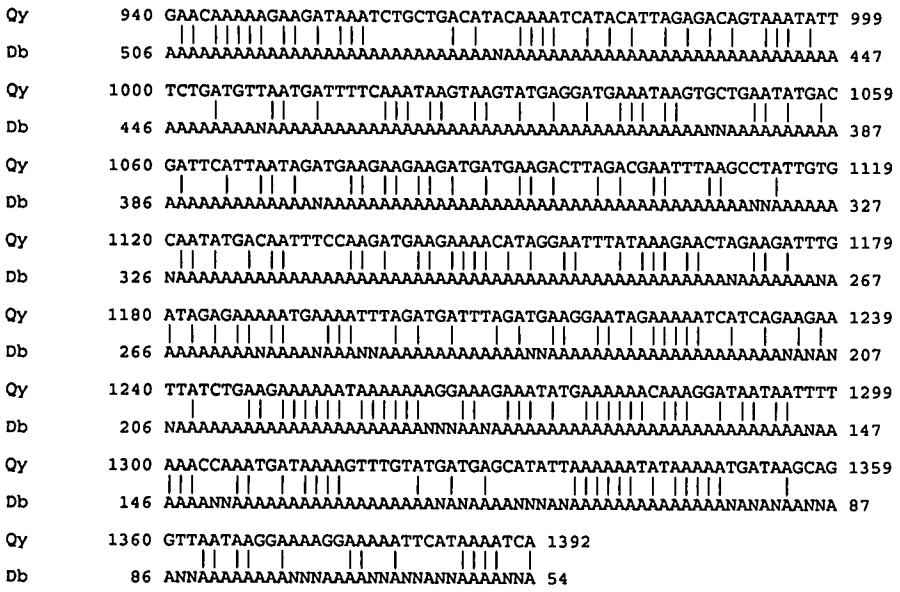
Query Match 15.7%; Score 232.2; DB 10; Length 1736;
Best Local Similarity 47.8%; Pred. No. 1.9e-28;
Matches 645; Conservative 0; Mismatches 705; Indels 0; Gaps 0;





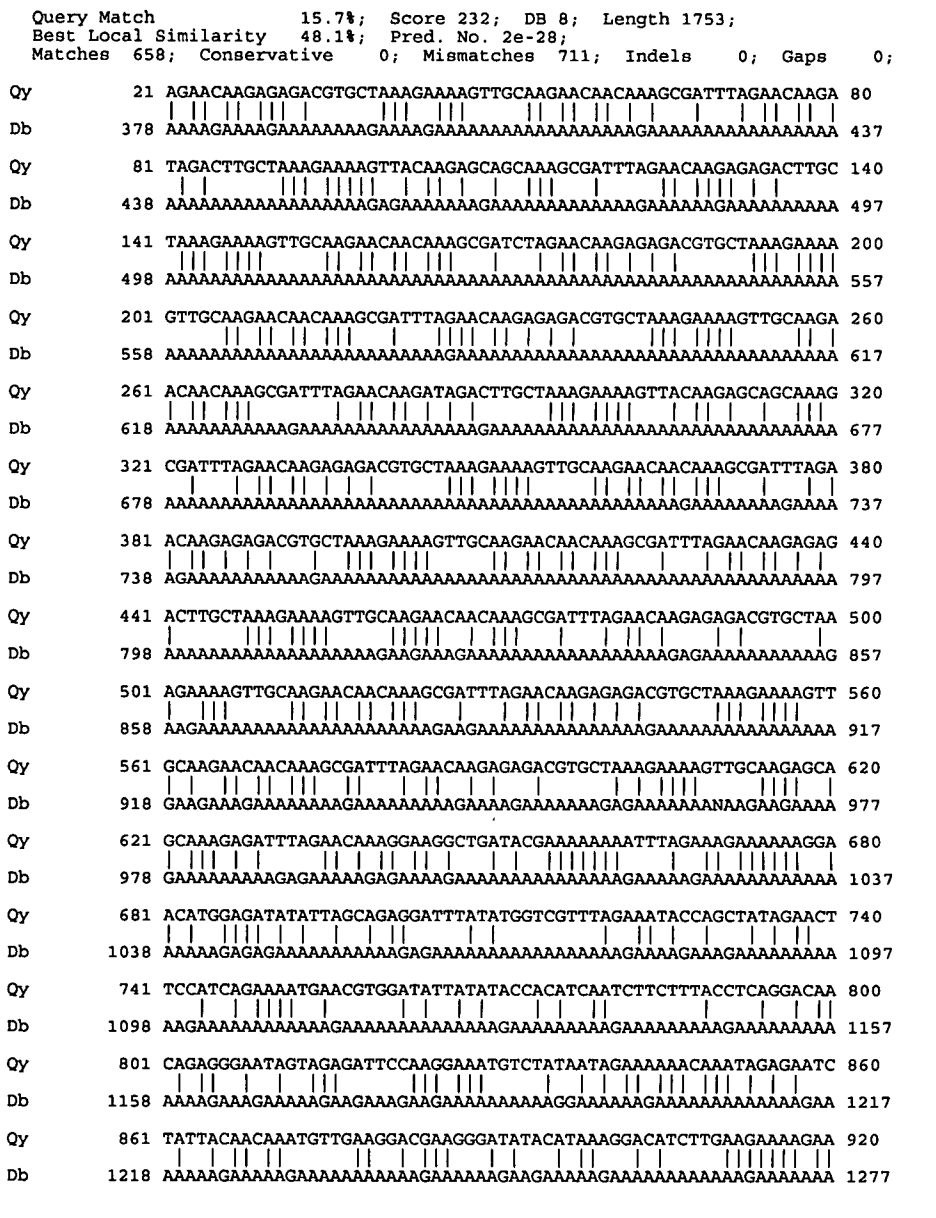
RESULT 18
AJ592026/c
LOCUS AJ592026 1407 bp DNA linear GSS 15-JAN-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 600D05, genomic survey sequence.
ACCESSION AJ592026
VERSION AJ592026.1 GI:37941650
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1407)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).
FEATURES Location/Qualifiers
source 1. .1407
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="600D05"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Wassilewskija"
misc_feature 1. .1407
/note="T-DNA flanking sequence
right border"
ORIGIN

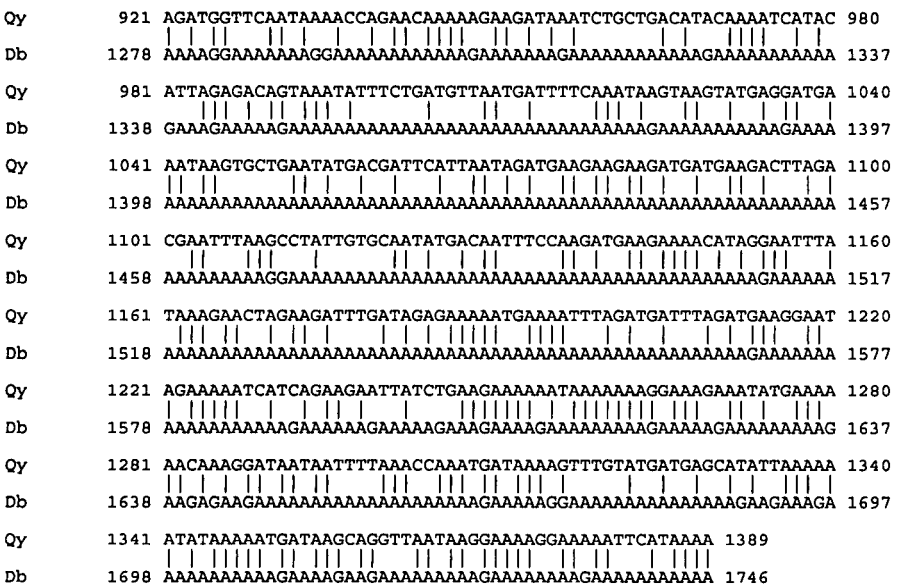




RESULT 19
DR131208
LOCUS DR131208 1753 bp mRNA linear EST 16-JUN-2005
DEFINITION 49217594 Drosophila pseudoobscura embryonic cDNA library Drosophila pseudoobscura cDNA clone E9 5', mRNA sequence.
ACCESSION DR131208
VERSION DR131208.1 GI:67876120
KEYWORDS EST.
SOURCE Drosophila pseudoobscura
ORGANISM Drosophila pseudoobscura
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1753)
AUTHORS Richards,S., Liu,Y., Bettencourt,B.R., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M.J., Chen,R., Meisel,R.P., Couronne,O., Hua,S., Smith,M.A., Zhang,P., Liu,J., Bussemaker,H.J., van Batenburg,M.F., Howells,S.L., Scherer,S.E., Sodergren,E., Matthews,B.B., Crosby,M.A., Schroeder,A.J., Ortiz-Barrientos,D., Rives,C.M., Metzker,M.L., Muzny,D.M., Scott,G., Steffen,D., Wheeler,D.A., Worley,K.C., Havlak,P., Durbin,K.J., Egan,A., Gill,R., Hume,J., Morgan,M.B., Miner,G., Hamilton,C., Huang,Y., Waldron,L., Verduzco,D., Clerc-Blankenburg,K.P., Dubchak,I., Noor,M.A., Anderson,W., White,K.P., Clark,A.G., Schaeffer,S.W., Gelbart,W., Weinstock,G.M. and Gibbs,R.A.
TITLE Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution
JOURNAL Genome Res. 15 (1), 1-18 (2005)
PUBMED 15632085
COMMENT Contact: Stephen Richards
Human Genome Sequencing Center
Baylor College of Medicine
One Baylor Plaza, Houston, TX 77030, USA
Tel: 713-798-6667
Email: stephenr@bcm.tmc.edu
NCBI Trace Archive: 226709361
Insert Length: 1750 Std Error: 0.25.
FEATURES Location/Qualifiers
source 1. .1753
/organism="Drosophila pseudoobscura"
/mol_type="mRNA"
/db_xref="taxon:7237"
/clone="E9"
/dev_stage="0-18h embryos"
/clone_lib="Drosophila pseudoobscura embryonic cDNA library"
/note="Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; oligo dt priming from poly A+ RNA, directionally cloned"
ORIGIN

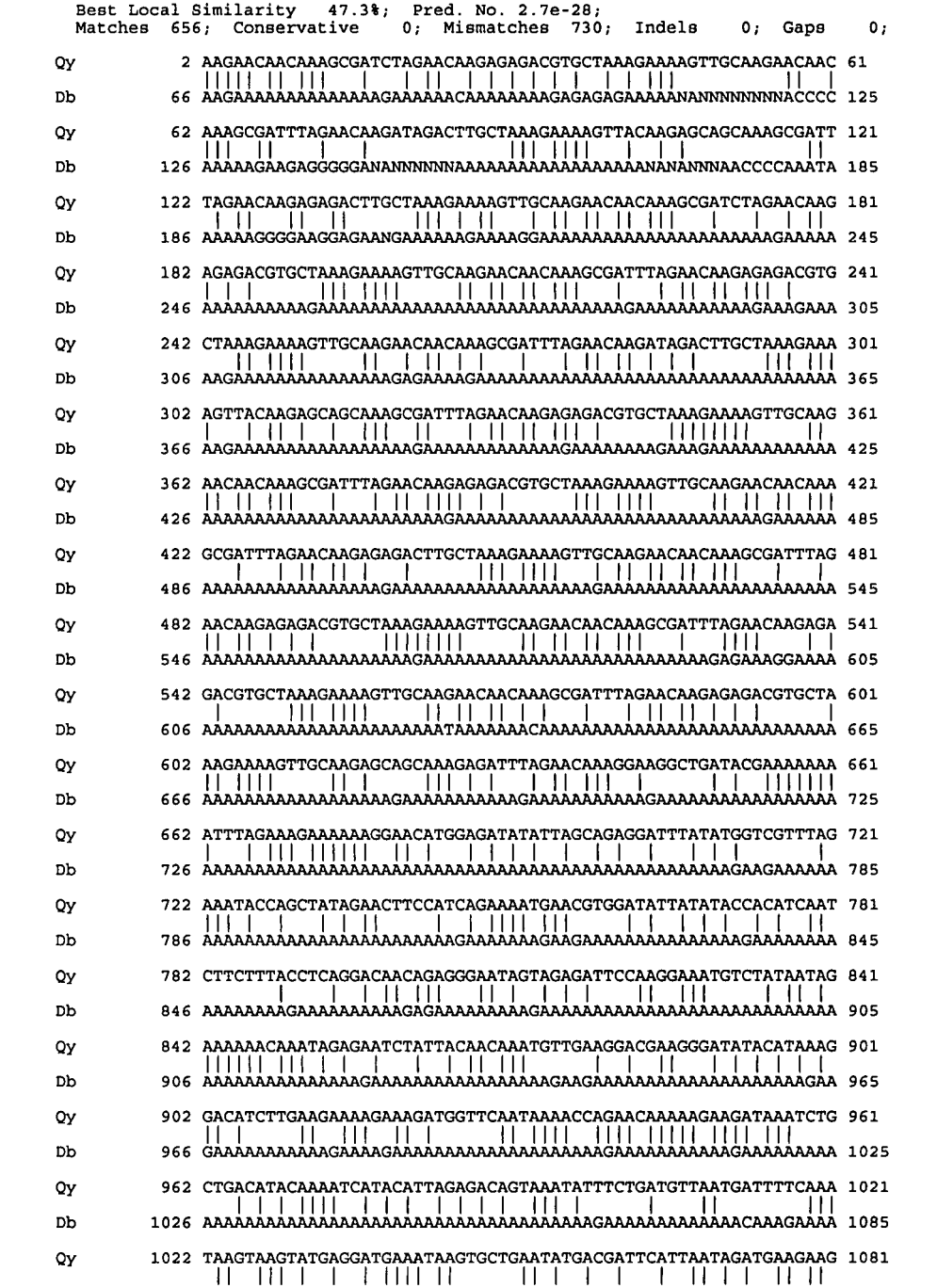
Query Match 15.7%; Score 232; DB 8; Length 1753;
Best Local Similarity 48.1%; Pred. No. 2e-28;
Matches 658; Conservative 0; Mismatches 711; Indels 0; Gaps 0;

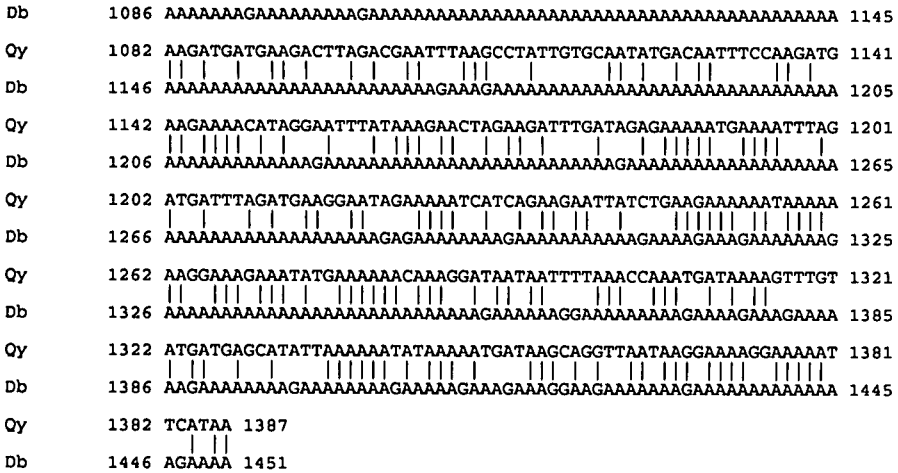




RESULT 20
CL113943
LOCUS CL113943 1512 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-59E24_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59E24, genomic survey sequence.
ACCESSION CL113943
VERSION CL113943.1 GI:40607578
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 880
High quality sequence stop: 1356.
FEATURES Location/Qualifiers
source 1. .1512
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-59E24"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

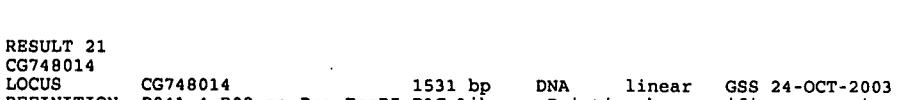
Query Match 15.6%; Score 231.2; DB 10; Length 1512;

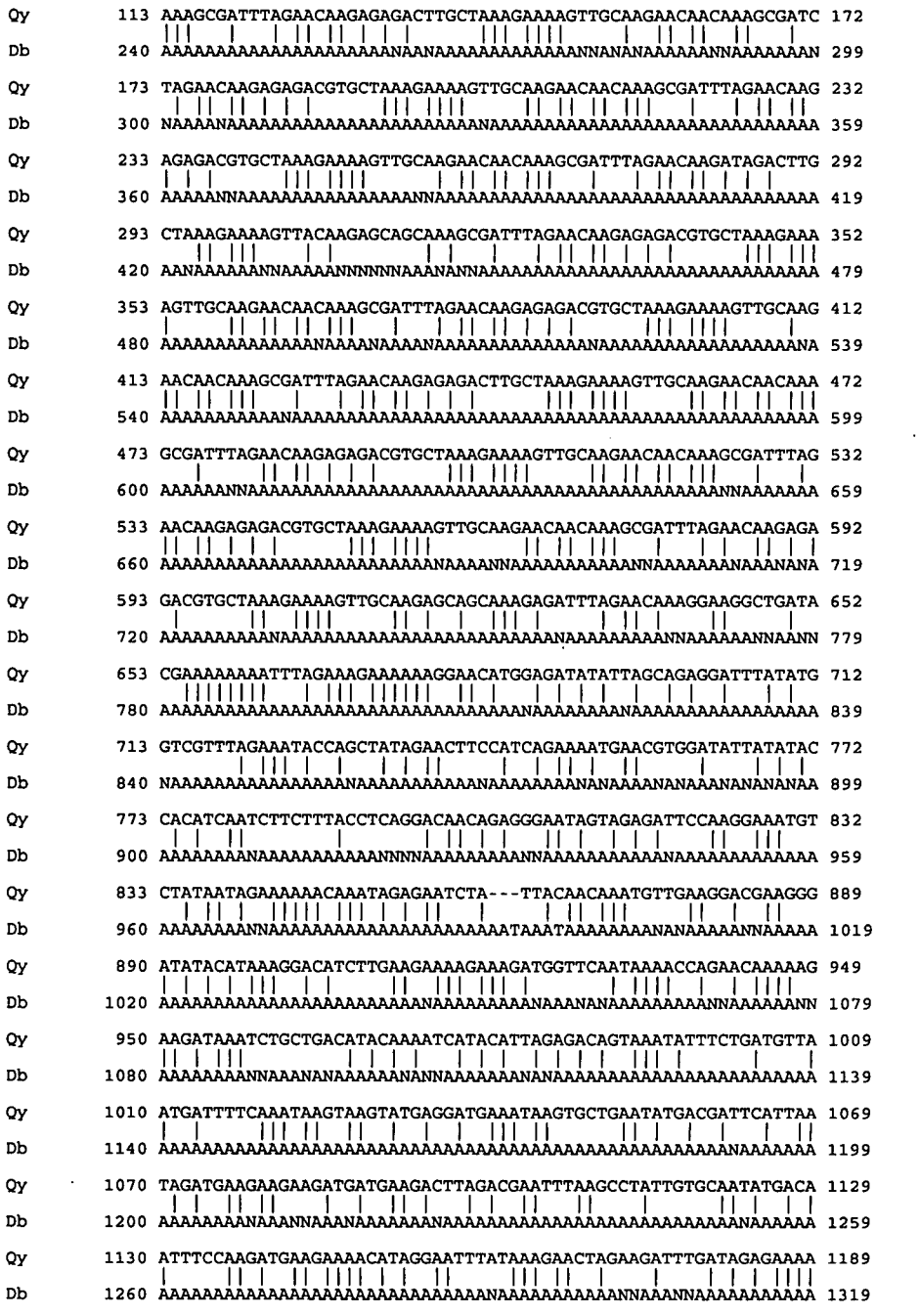


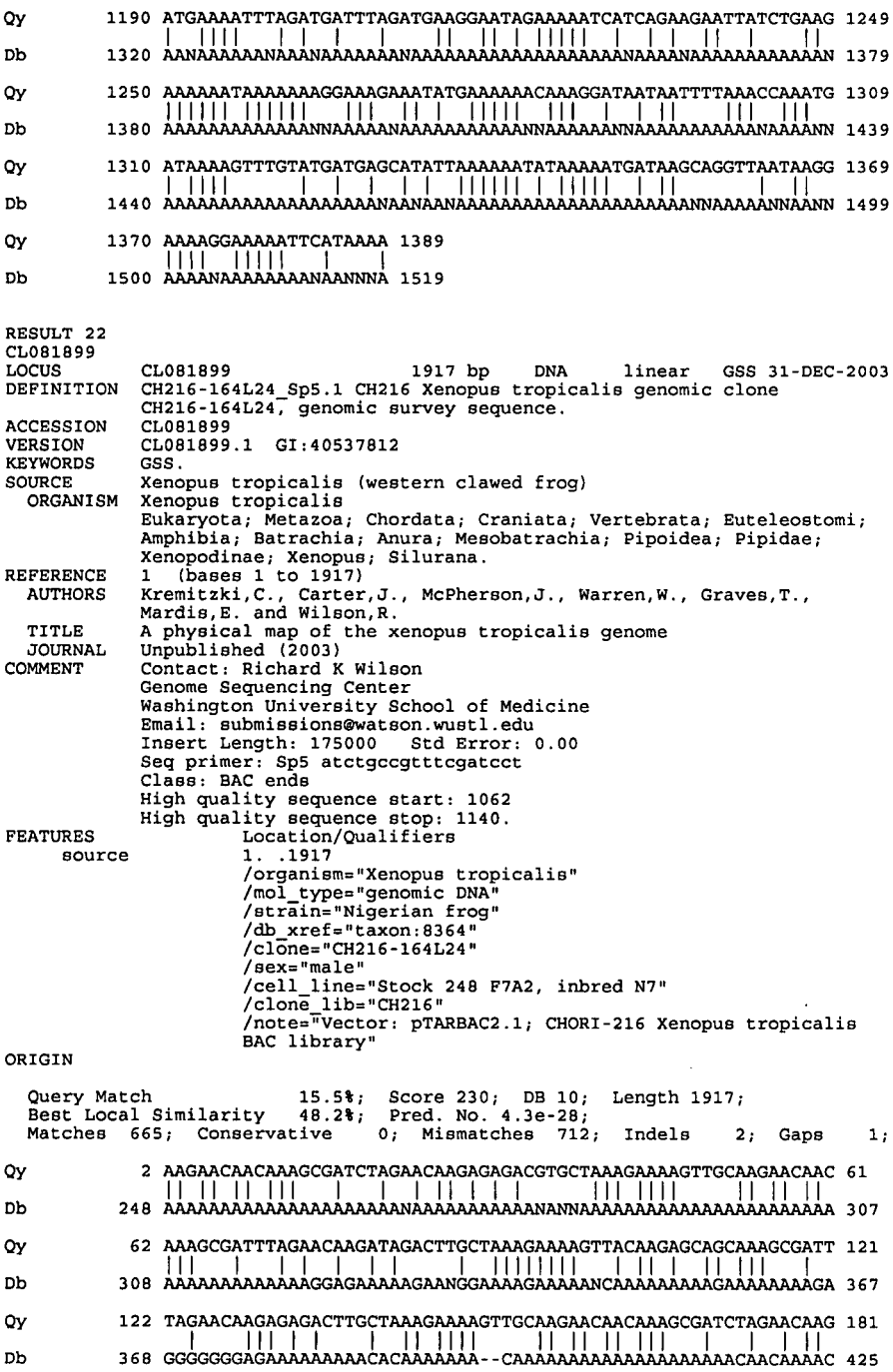


RESULT 21
CG748014
LOCUS CG748014 1531 bp DNA linear GSS 24-OCT-2003
DEFINITION P041-4-B02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG748014
VERSION CG748014.1 GI:37968940
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1531)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1531
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 15.6%; Score 231.2; DB 10; Length 1531;
Best Local Similarity 44.3%; Pred. No. 2.7e-28;
Matches 594; Conservative 0; Mismatches 743; Indels 3; Gaps 1;







```
Qy       1190 ATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAG 1249
Db       1320 AANAAAAAANAAANAAAAAAANAAAAAAAAAAAAAAAAAAAAANAAAANAAAAAAAAAAN 1379

Qy       1250 AAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATG 1309
Db       1380 AAAAAAAAAAAAAANNAAAAANAAAAAAAAAAAANNAAAAAANNAAAAAAAAAAAANAAANN 1439

Qy       1310 ATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGG 1369
Db       1440 AAAAAAAAAAAAAAAAAAANAANAANAAAAAAAAAAAAAAAAAAAAANNAAAAANNAANN 1499

Qy       1370 AAAAGGAAAAATTCATAAAA 1389
Db       1500 AAAANAAAAAAAANAANNNA 1519


RESULT 22
CL081899
LOCUS       CL081899                1917 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-164L24_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-164L24, genomic survey sequence.
ACCESSION   CL081899
VERSION     CL081899.1  GI:40537812
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1917)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1140.
FEATURES             Location/Qualifiers
     source          1. .1917
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-164L24"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.5%;  Score 230;  DB 10;  Length 1917;
  Best Local Similarity   48.2%;  Pred. No. 4.3e-28;
  Matches  665;  Conservative    0;  Mismatches  712;  Indels    2;  Gaps    1;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db        248 AAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAANANNAAAAAAAAAAAAAAAAAAAAAA 307

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db        308 AAAAAAAAAAAAAAGGAGAAAAAGAANGGAAAAGAAAAANCAAAAAAAAAGAAAAAAAAGA 367

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db        368 GGGGGGGAGAAAAAAAAACACAAAAAAA--CAAAAAAAAAAAAAAAAAAAACAACAAAAC 425

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db        426 ACAAACAAAAAAACAAAAAAAAAAAAAACAAAACAAAGAAAAGANAAAAAAGAAAAAAGAA 485

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db        486 GAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAAAAAGAAAAA 545

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db        546 AAAAAAAGAAAAAGAAAAAAAAAAAAAAAGAAAAAAAAAAAAACAAAAAAAAAAAGAAAAA 605

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db        606 AAAAAAAAAAAAAAAAAGAAAAAGAAAAAAGAAAAAAAAAAGGAGAAAAAAAAAAAAAGAAA 665

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db        666 AAAAAAAAAAAGAAAAAAAAGAAAAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 725

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db        726 AAGAAAAAAGAAAAAAAAGAGAAAAAGAAAAAGGAGAAAAAAAAAAAAACAAAAAAAAAAG 785

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db        786 AAAAAAAAAAGAAAAAGAAAAAAAAAAACAGAAAAAAAAAAAAAAGAGAAAAAAAAAAAA 845

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db        846 AAGACAGAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAGAGAAAAAAAGAAAGAAAA 905

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db        906 AGAAAAAAAAAAGAAAGAATAAAAGAAAAAAAAAAAAAAAAAAAAAAAAGAAAGAAAAAA 965

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db        966 AAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAGAAAAAA 1025

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
Db       1026 AAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAGAAAAAAA 1085

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db       1086 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAA 1145

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db       1146 AAAAATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAGAAAAAAAAAAAAAAAAAAAG 1205

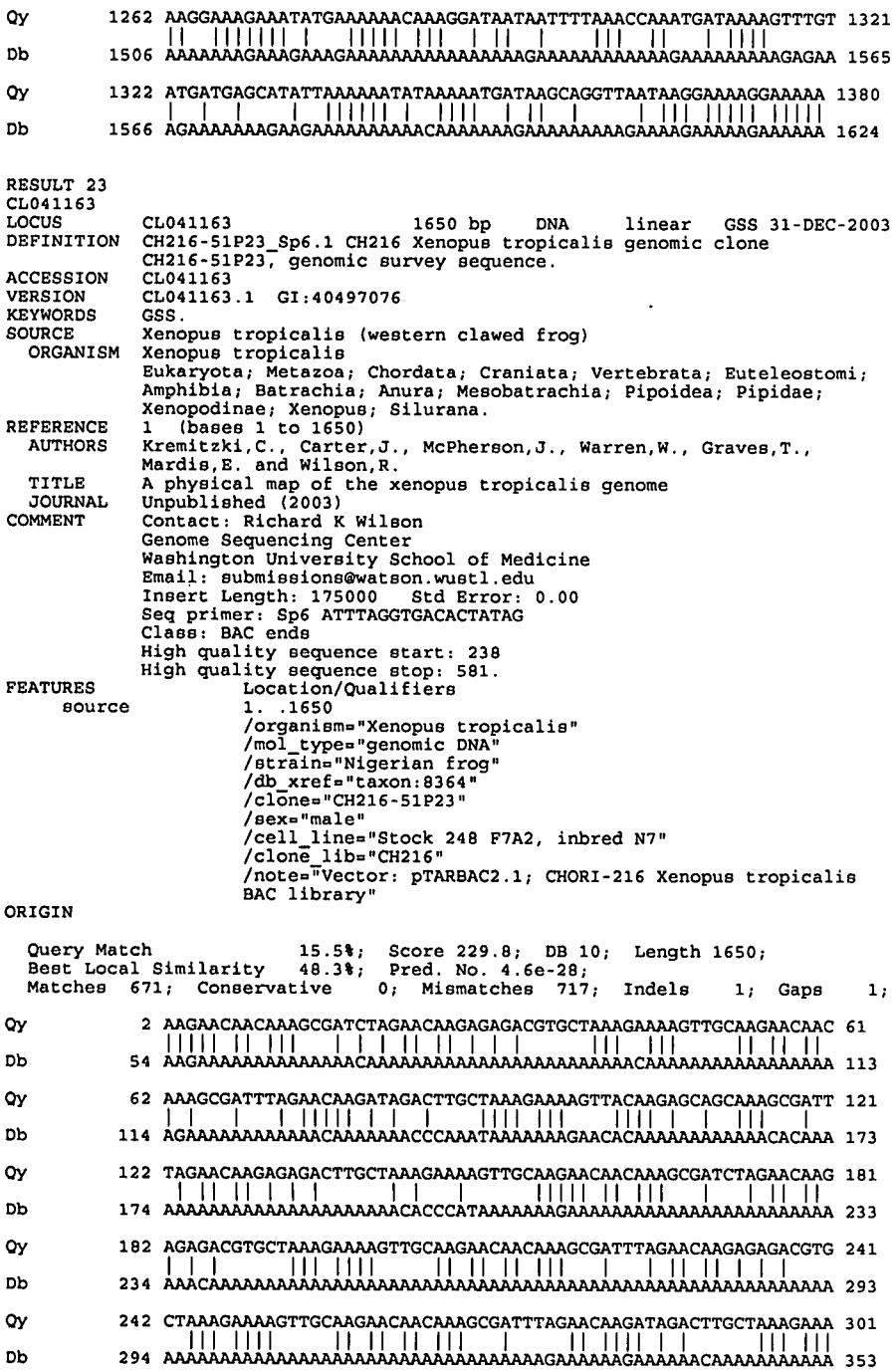
Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       1206 ACAAAGAAAAAAAAAAAGAAAGAAAAAAAAGAAAAAAAGAAAAAAGAAAAAAAAAAAAAAA 1265

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db       1266 AAAAAAAGAAAGAAAAGAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAAAAAAGGAAAA 1325

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
Db       1326 AAGAAAAAGAAAAAAAAAGCAAAGAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1385

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
Db       1386 AAAAAAAAAAAAGAAAAAAAAAGAAAGAAAAAAAAAAAAAAAAGAAGAAAAAAAAAAGAAG 1445

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db       1446 AAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAACAAAAAAAGAAAAAAAAAAAAAAA 1505
```



```
Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
Db       1506 AAAAAAAGAAAGAAAGAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAGAAAAAAAAAGAGAA 1565

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1566 AGAAAAAAAGAAGAAAAAAAAAACAAAAAAAGAAAAAAAAAGAAAAGAAAAAGAAAAAA 1624


RESULT 23
CL041163
LOCUS       CL041163                1650 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-51P23_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-51P23, genomic survey sequence.
ACCESSION   CL041163
VERSION     CL041163.1  GI:40497076
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 238
            High quality sequence stop: 581.
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-51P23"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.5%;  Score 229.8;  DB 10;  Length 1650;
  Best Local Similarity   48.3%;  Pred. No. 4.6e-28;
  Matches  671;  Conservative    0;  Mismatches  717;  Indels    1;  Gaps    1;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db         54 AAGAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAA 113

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db        114 AGAAAAAAAAAAAAACAAAAAAACCCAAATAAAAAAAGAACACAAAAAAAAAAAACACAAA 173

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db        174 AAAAAAAAAAAAAAAAAAAAACACCCATAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAA 233

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db        234 AAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 293

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db        294 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAACAAAAAAAAAA 353

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db        354 AAAAAAGAGAAGAGAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAA 413

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db        414 AAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 473

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db        474 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC 533

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db        534 AAAAAAAAAAAAAAAAAAGAAAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAACAA 593

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db        594 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAA 653

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db        654 AACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 713

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db        714 ACAAAAGAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAA 773

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db        774 AAAAAAAAAAAAAAATAAAAAAAAAAAACAAAATAAAAAAAAAAAAAAAACAAAAAAAAAA 833

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
Db        834 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 893

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db        894 AAAAAAACAAAAAAGAAAAAGAAAAAAAAAAAAAAAGAAAACAAAAAAAAAAAAAAAAAA 953

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db        954 AAGAAAAAAAAAAACAAAAAAACAAAAAAAAAACAAAAACAAAAAAGAAAAAAAAAAAAA 1013

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       1014 AAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1073

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db       1074 AAACAAAAAAAAAAAAAAAAAAAAAAAAAACAAGAAGAAAAAACAAAAAAAAAAAAAAAA 1133

Qy       1082 AAGATGATGAAGACTTAGAC-GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
Db       1134 AAAAAAAAAGAAAAAGAAAGAACAAAAAACAAAACAAAAAAGAGAACAAAAAAAAAAA 1193

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
Db       1194 GAAAAAACGACAAACAAAAAAAAAAAAAAAAAAGAAAAAAAGAACAAAAAAAAAAAAA 1253

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
Db       1254 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAGAAAAAAAAAAAAAAAAAA 1313

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
Db       1314 AAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAGACAAAAAAACA 1373

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1374 AAAAACAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAACAAAGAAAACAAAAAAAAAAA 1433
```

```
Qy       1381 TTCATAAAA 1389
              | ||||
Db       1434 AAAAAAAAA 1442


RESULT 24
CL082587
LOCUS       CL082587               1762 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-168C5_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-168C5, genomic survey sequence.
ACCESSION   CL082587
VERSION     CL082587.1  GI:40538500
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1762)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 1020
            High quality sequence stop: 1162.
FEATURES             Location/Qualifiers
     source          1. .1762
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-168C5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.5%;  Score 229.6;  DB 10;  Length 1762;
  Best Local Similarity   47.9%;  Pred. No. 5e-28;
  Matches  712;  Conservative    0;  Mismatches  767;  Indels    7;  Gaps    2;

Qy           2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
               || || || |||   |   | || ||   | |   |  | || ||      || || ||
Db         269 AAAAAAAAAAAAAAAAANAAAAAAAAAANNAAAAAAGAGAGAGGAAGAAAAAAAAAAAAA 328

Qy          62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
               |||  ||   | || || | | |       | |  |||   | || |  |  |||  ||
Db         329 AAAAAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAGAAAAAAAAAAAAAAAAGAAA 388

Qy         122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
                | || || | | |       | ||||||    ||||| || |||   |   | || ||
Db         389 AAAAAAAAAAAAAAAAAAAAAGAGGAAAAGAAAAAAGAAAAAAAAAAAAAAGAGAAAAAAAA 448

Qy         182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
               | | |      ||| ||||     || || || |||  ||   | || |||| | |
Db         449 AAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAAAGAAAAAAAAA 508

Qy         242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
                |||| ||     || || |  |||       | || ||   | |       ||| ||
Db         509 AAAAAGAGAAAAAAAAAAAAAAAGAAAAAAAGAAAAAAAGAAAGAAAAAAAANGAAAAAAG 568

Qy         302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
               |   | || |  || |||   |  | || || | | |     ||| ||||      |||
Db         569 AAAAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 628

Qy         362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
               || ||  ||| ||   | || || | |||       ||||||      | ||  || ||
Db         629 AAAAAAGAAGAGAAAAAAAAAAAAAAAAAGAAAAANAGGAGAAAAAAAAAGAAAAACACCAA 688

Qy         422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
                 |   | ||  | |||| |      ||  ||||   | || || || |||   |   |
Db         689 AAAAAAGAAAAAGAAAGAGGCAAAAAAAGAAAAAAAGGAAAAAAAAAAAAAAAACAAAAAA 748

Qy         482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
               ||     |           ||| |||       || ||  | | |  ||    | || || | |
Db         749 AAAGGAAAGAGAAAAAAAAAAAAAGAAAAAAAAAAAGAAAGAAAGAAAAAAAAGAAAAAA 808

Qy         542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
               |      ||| |||||  | |  ||  | |||    |   ||||  || | |||
Db         809 AAAAAAAGAAAAAAAAGAAGAAGAAAAGAAAAAAAAAGAAAGAAAAAAAAAGAAAAAAAG 868

Qy         602 AAGAAAAGTTGC------AAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGA 655
               || |||||          || || |  ||| | |    | || |||| ||| |  | |  |
Db         869 AAAAAAAGCAAAGAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAGAAAGACAAAAAAAA 928

Qy         656 AAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTC 715
               |||||||    | ||||||||||  || |      | | |  || | |   |      |
Db         929 AAAAAAAAGAAAAAAGAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAGAAAAAAA 988

Qy         716 GTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCAC 775
                 | ||| |  ||     | ||       |  | ||||  ||    | |  | | |  |
Db         989 AGAGAAAAAGAAAAGAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAAAA 1048

Qy         776 ATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTA 835
               |  ||         |    || | || |||   || |  | | |     ||  |||     |
Db        1049 AAAAAGAAAAAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1108

Qy         836 TAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATAC 895
                || | |||||    || | | ||    |  | || |||      ||  |  ||    | | |
Db        1109 AAAAAAAAAAAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAA 1168

Qy         896 ATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATA 955
               | ||| || |    || |||| ||| |  |   || ||||  | || |||||    ||
Db        1169 AAAAAAGAGAAGAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAACGGAAG 1228

Qy         956 AATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATT 1015
               ||       | | |||||  |  |  | | | |  |||       || |  ||  |
Db        1229 AAAAGAAGAAGAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAAAGAAGAAAAAAAAG 1288

Qy        1016 TTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATG 1075
                ||| ||  ||  |   | |  ||| ||      ||  |  |  |   |  || | |
Db        1289 AAAAAAAAAAGAAAAAGAGGAAAAAAAAAAAAAAAAGAGAAAAAAAAAAAAAAAAAGAAAAA 1348

Qy        1076 AAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCC 1135
               || || ||||  |  || |  | |  ||  ||    |      || |  | ||
Db        1349 AAAAAAAAGAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAAAAAAACAGA 1408

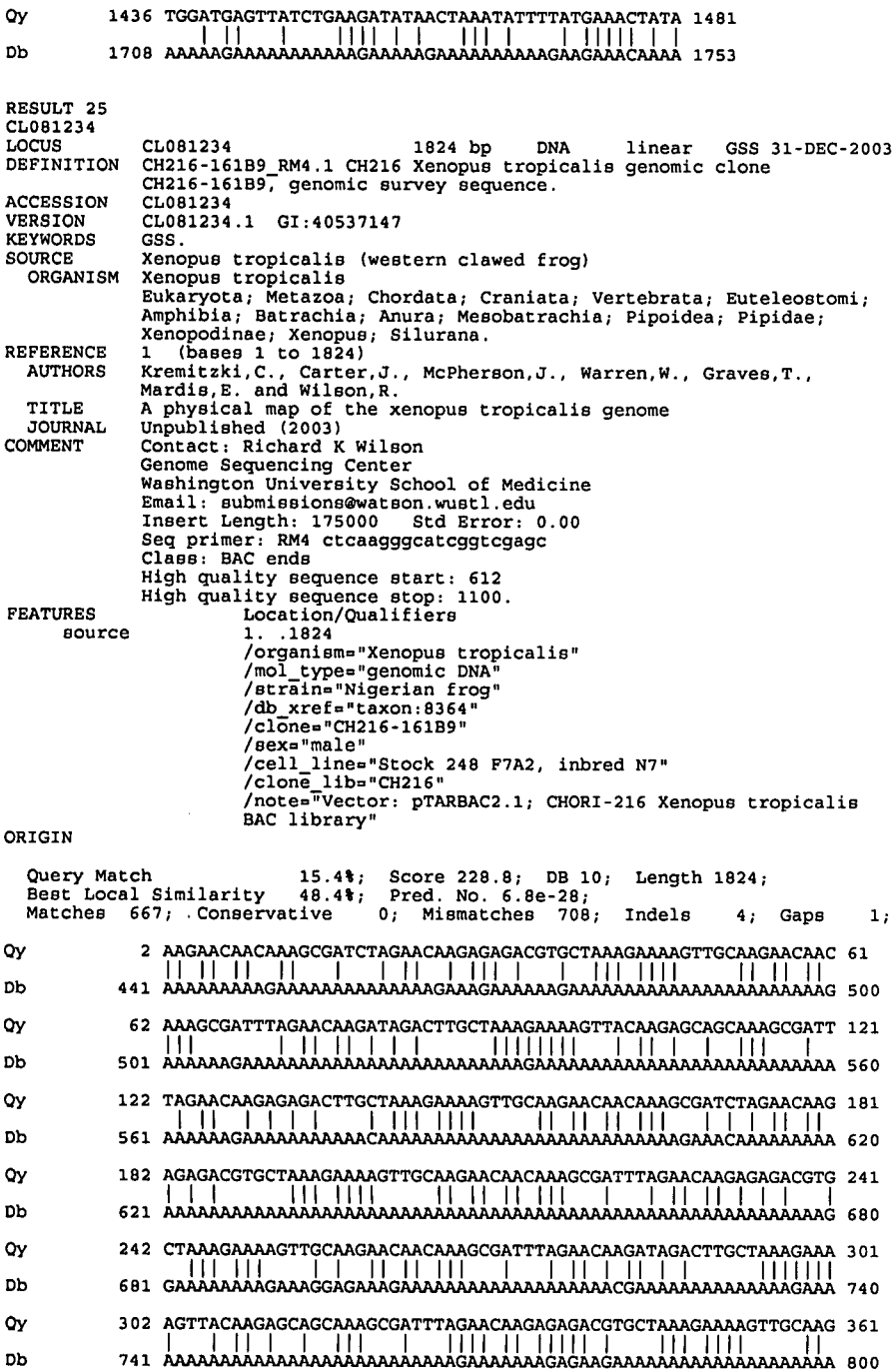
Qy        1136 AAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAA 1195
               || |  |||||| | |  ||   |  || ||  |||| |    | | | |||||  |||
Db        1409 AAAAGAAAGAAAAAAAAAAAAAAAAAGAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAA 1468

Qy        1196 ATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAA 1255
               |  | | ||   | |  || ||| | | |||  |  | ||||| |    || ||||||
Db        1469 ACGAAAAAGAAAAAAAAAAAAGAA-AAAGAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA 1527

Qy        1256 TAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAA 1315
               ||| |||  ||||||| | || ||| | | || || ||    | |  ||    ||||
Db        1528 AAAACAAAAAAAAGAAAGAAAAAGAAAAAGAAGAGAAAAAAAAAAGAAAGAAAAGAAAAA 1587

Qy        1316 GTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGG 1375
               |  | |  |  | |    ||||| | | |||  | ||  |    || |  |||||
Db        1588 GAAAGAAAAAAAAAAAACAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAAAAAGAAAAAGA 1647

Qy        1376 AAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCG 1435
               |||||   | |||| |       | | |    |  |||| || |||     | |
Db        1648 AAAAAAAAAGAAAAAAAGAAAAAAAAAAAAAAAAGAGAGAAAGAAGAAAGAAAAGAAA 1707
```



```
Qy       1436 TGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTATA 1481
Db       1708 AAAAAGAAAAAAAAAAAGAAAAAGAAAAAAAAAAAGAAGAAACAAAA 1753

RESULT 25
CL081234
LOCUS       CL081234                1824 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-161B9_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-161B9, genomic survey sequence.
ACCESSION   CL081234
VERSION     CL081234.1  GI:40537147
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1824)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 612
            High quality sequence stop: 1100.
FEATURES             Location/Qualifiers
     source          1. .1824
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-161B9"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.4%;  Score 228.8;  DB 10;  Length 1824;
  Best Local Similarity   48.4%;  Pred. No. 6.8e-28;
  Matches  667;  Conservative    0;  Mismatches  708;  Indels    4;  Gaps    1;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
Db        441 AAAAAAAAAGAAAAAAAAAAAAAAGAAAGAAAAAAGAAAAAAAAAAAAAAAAAAAAAAG 500

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
Db        501 AAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAA 560

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
Db        561 AAAAAAGAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAACAAAAAAAA 620

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
Db        621 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 680

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
Db        681 GAAAAAAAAGAAAGGAGAAAGAAAAAAAAAAAAAAAAAACGAAAAAAAAAAAAAAAGAAA 740

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
Db        741 AAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAGAGAAGAAAAAAAAAAAAAAAAAA 800

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
Db        801 AAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAA 860

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
Db        861 AAAAAAAACAAAAAAAAAAAAAAAAAAAAAAGAAAAAGAGAAAAAAAAAAAAAAAAAAAA 920

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
Db        921 AAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAACA 980

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
Db        981 AAAAAAAAAAAAGAAAAAAGAAAAAAAAGAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAA 1040

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
Db       1041 AAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAAAA 1100

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
Db       1101 AAGGAAAAAAAAAGAAGAAAAAAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAGAAAAAAA 1160

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
Db       1161 AAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAA 1220

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
Db       1221 AAAGAGAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAGAA 1280

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
Db       1281 AAAAGAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAAAA 1340

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
Db       1341 AAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAA 1400

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
Db       1401 AAAAAAAAAAAAAGGAAAAGAAAAAAAAAAGAAAAAAAAAAACAAGAAAAAAAAA----AAA 1456

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
Db       1457 AAGAAAAAAAAAAAGAACAAAAGAAGAAAAAAAAAAAAAAGAGATAGAAAAAAAAAAAAAAAA 1516

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
Db       1517 AAAAAAAAAAAAAAGAGAGAAAAAAAAAAAAANGAAGAAAAAAAAAAAAAGAAAAGAAAAAAAA 1576

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
Db       1577 AAAAAAAAAAAAAAAGGAAAAGAAAGAGAAAAAGAAAAGACAGGAAAAAGAGAAAAAAAAG 1636

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
Db       1637 GATAAAAAAAAAAACAAAAAAGAGAGAAAAAAAAGAAAAAAAAGAAAAAAAAAAAGAAAA 1696

Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
Db       1697 AAAAAAAAAAAAAAAAAAAAAGAAGACAAAAACAAAAAAAAGGAAAAGAGAAAAAAAACA 1756

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db       1757 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAAGAAAAAAAAAAAAAAA 1815

RESULT 26
CL118815
LOCUS       CL118815                1727 bp    DNA     linear   GSS 05-JAN-2004
```

```
DEFINITION  ISB1-72O5_T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72O5,
            genomic survey sequence.
ACCESSION   CL118815
VERSION     CL118815.1  GI:40612450
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1727)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 903
            High quality sequence stop: 1314.
FEATURES             Location/Qualifiers
     source          1. .1727
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-72O5"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             15.4%;  Score 228.6;  DB 10;  Length 1727;
  Best Local Similarity   47.8%;  Pred. No. 7.3e-28;
  Matches  663;  Conservative    0;  Mismatches  724;  Indels    0;  Gaps    0;

Qy          3 AGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACA 62
              | || || ||  | ||  | || | | |   | ||| ||||     || || | |
Db        272 AAAAAAAAAACACAATAAAAAAAACAAAAAAAAAGGAAAAAAAAAAAAAAAAAAAAAAA 331

Qy         63 AAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTT 122
              |  | ||  | || | || | |       ||| ||||   | || |  |  |||
Db        332 ACACAATAAAAAAAATGAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAGAAA 391

Qy        123 AGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGA 182
                || || | | |    | ||| ||||     || || || |||   |   | || || |
Db        392 GAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAACAAAAAAAAAAACACAAAAAAAAAA 451

Qy        183 GAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGC 242
              || |||  ||| ||||     || || || |||   |   | || || | | |  |
Db        452 GAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA 511

Qy        243 TAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAA 302
              |||  ||||   | || |  ||||||   |      || || | | |      ||  ||||
Db        512 AAAAAAAAAAAAGAAAAATAAACAAAAAAAAAAATAAAAAACACAAAAAAAAAGAACAAAAA 571

Qy        303 GTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGA 362
                | || |  |  |||   |   | || || | | |      ||| ||||     || |
Db        572 AAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAACAAACAAAAAAAAAAAAAAAA 631

Qy        363 ACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAG 422
              | || |||   |   | || |||| | |      ||| || |     || || || |||
Db        632 AAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAACAAAAAAAAAGAAAA 691

Qy        423 CGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGA 482
              |   | | || | | |      ||| |||      ||  |  | |||  ||   | |
Db        692 AAAGAAAACAAAAAAAAAAAAAAACAAAAAAATAAAAAAAATAAGAAAAAAAGAAAAAAA 751

Qy        483 ACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAG 542
              |||| | | |   |  ||| ||||     || || |  |||   |   | || |||| |
Db        752 ACAAAAAAAAGAAGACAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAACAAAAAGAAAA 811

Qy        543 ACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAA 602
              |   | ||  ||||      || || || |||   |   | ||  ||| |         ||
Db        812 AGAAGAAAAGAAAAAAAAAAAACAATAAAAAAAAAAAAAAAAAAACAGAAAACAAAAAAAA 871

Qy        603 AGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAA 662
              | ||||   | || |  | | | | |   | || |||  ||      | |  ||||||||
Db        872 AAAAAAAAAGAAAAAAAAAAACAAAAACAGAAAAAAAAAAAAAAAAAAAAAACAAAAAAAA 931

Qy        663 TTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGA 722
                | ||| ||||||  || |  | | | |  | ||| |  | | |       | |
Db        932 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAA 991

Qy        723 AATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATC 782
              || |  | | | | ||      |  | ||||  ||      | |  | | | |  | ||
Db        992 AAGAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1051

Qy        783 TTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGA 842
                  |  | || | |  ||     | | |      |  |||     | || | |
Db       1052 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAACAAAAAAAAAAAAAAAAA 1111

Qy        843 AAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGG 902
              ||||| ||| | | ||   |  | || |||     ||  |  ||   | | | | ||| |
Db       1112 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 1171

Qy        903 ACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGC 962
              | |    ||  ||| |||        || ||||  | || ||||| ||||  ||
Db       1172 AAAAAAAAAAACAAAAAAAAGAAAGAAAAAAAAAAAAAAAGAAAAAAAGAAGAAAAACA 1231

Qy        963 TGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAAT 1022
              || | | |||| | | |  | | | | ||| |       |    ||  |   |||
Db       1232 AGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1291

Qy       1023 AAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGA 1082
              ||  ||  |  |  |  ||| ||     || |  |  |   |  ||   |  || || |
Db       1292 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAA 1351

Qy       1083 AGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGA 1142
                |  |  | |  | | ||  ||  ||      ||  | ||     ||   |
Db       1352 GAAAAAAAGACAAAAAAAGAAAAAAAACAATAAAAAAAAAAGAAAAAAAAAAAAAATAAA 1411

Qy       1143 AGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGA 1202
              | |||| | |  ||  | ||| |  | || |       | | ||||| ||||   | |
Db       1412 AAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAACA 1471

Qy       1203 TGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAA 1262
                |  | |  ||  || | ||||   |  | || ||  |   || ||||||  || |||
Db       1472 AAAAACAAAAAAAAAAAAAAAAAAGAAAAAATAAAAAAAAAACAAAAAAAAAAAACAAA 1531

Qy       1263 AGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTA 1322
              |   || ||| |  |||||| |||  | || ||    |||| |||  | ||||    | |
Db       1532 AAACAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAACGAAAAAAAAAAAAAAAGAA 1591

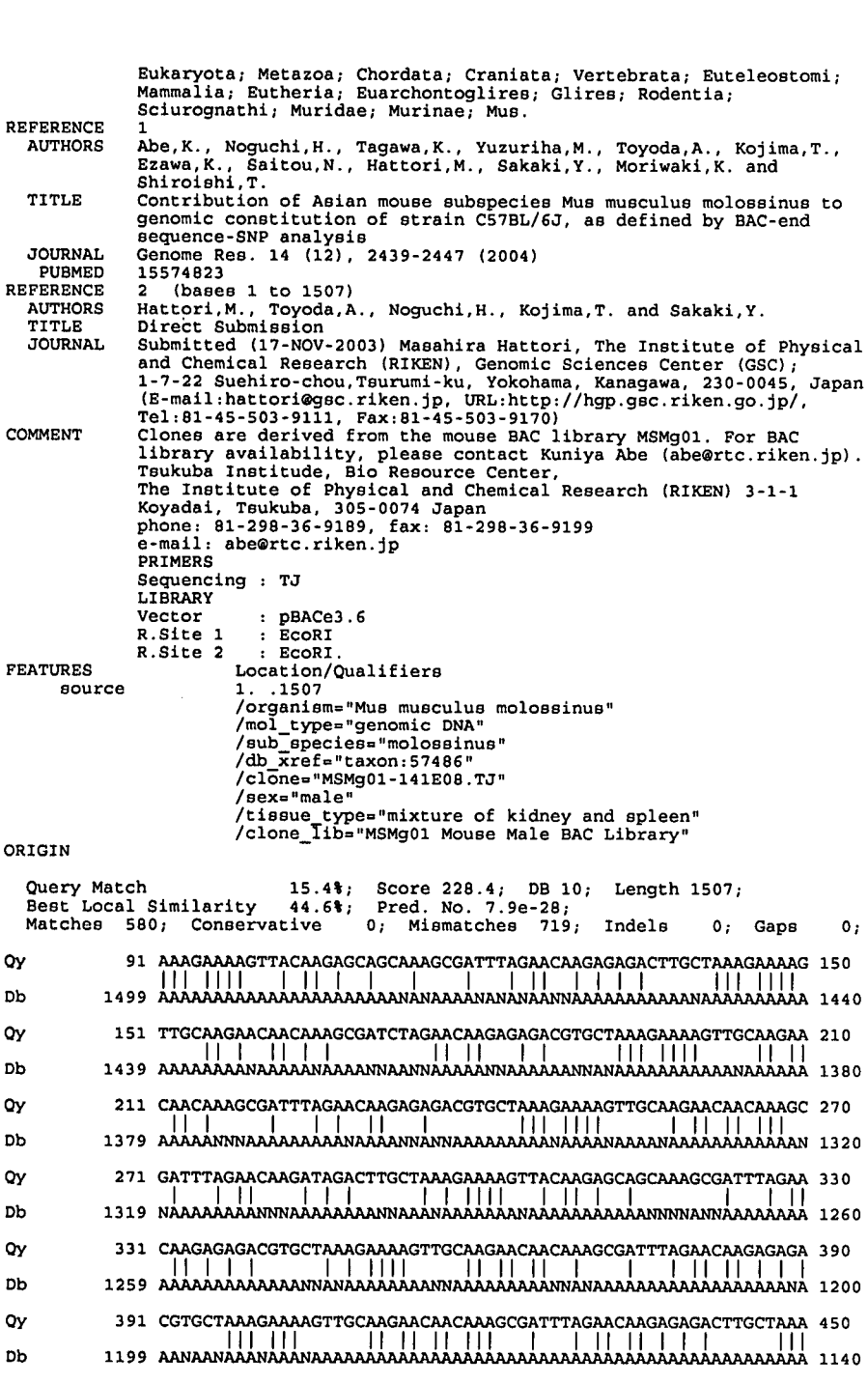
Qy       1323 TGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATT 1382
                |  | |  ||||| | ||||| || ||  |     ||  |  |||     | ||
Db       1592 GAAAAAAAAAGAAAAAAAAAAAAAAAGAAAAAAAAAACAAATAAAAAAGAACATAAAA 1651

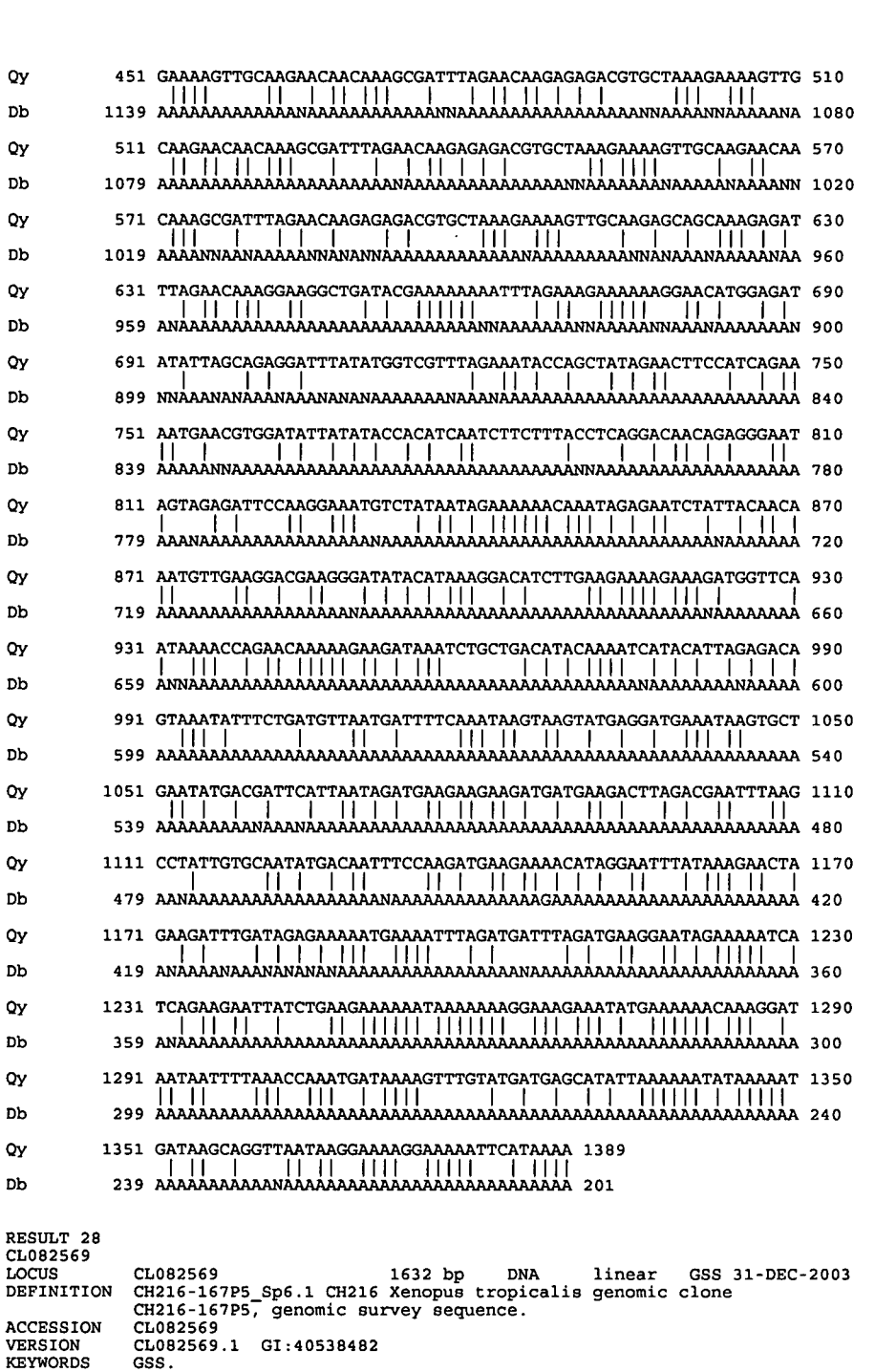
Qy       1383 CATAAAA 1389
                | ||||
Db       1652 CGAAAAA 1658


RESULT 27
AG346189/c
LOCUS       AG346189               1507 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-141E08.TJ, genomic survey
            sequence.
ACCESSION   AG346189
VERSION     AG346189.1  GI:47919499
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1507)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1507
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-141E08.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match           15.4%;  Score 228.4;  DB 10;  Length 1507;
  Best Local Similarity   44.6%;  Pred. No. 7.9e-28;
  Matches  580;  Conservative    0;  Mismatches  719;  Indels    0;  Gaps    0;
```

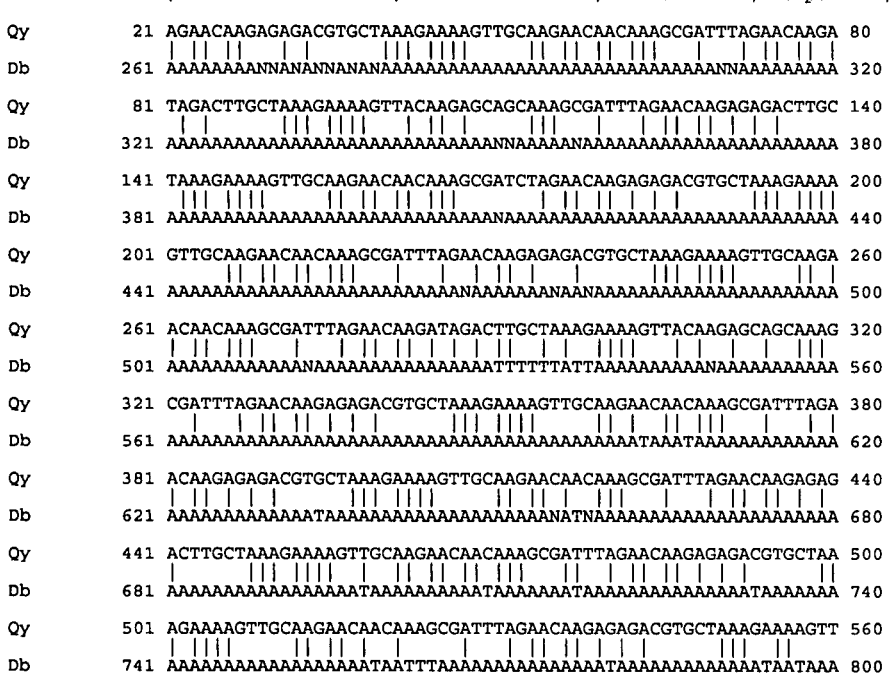


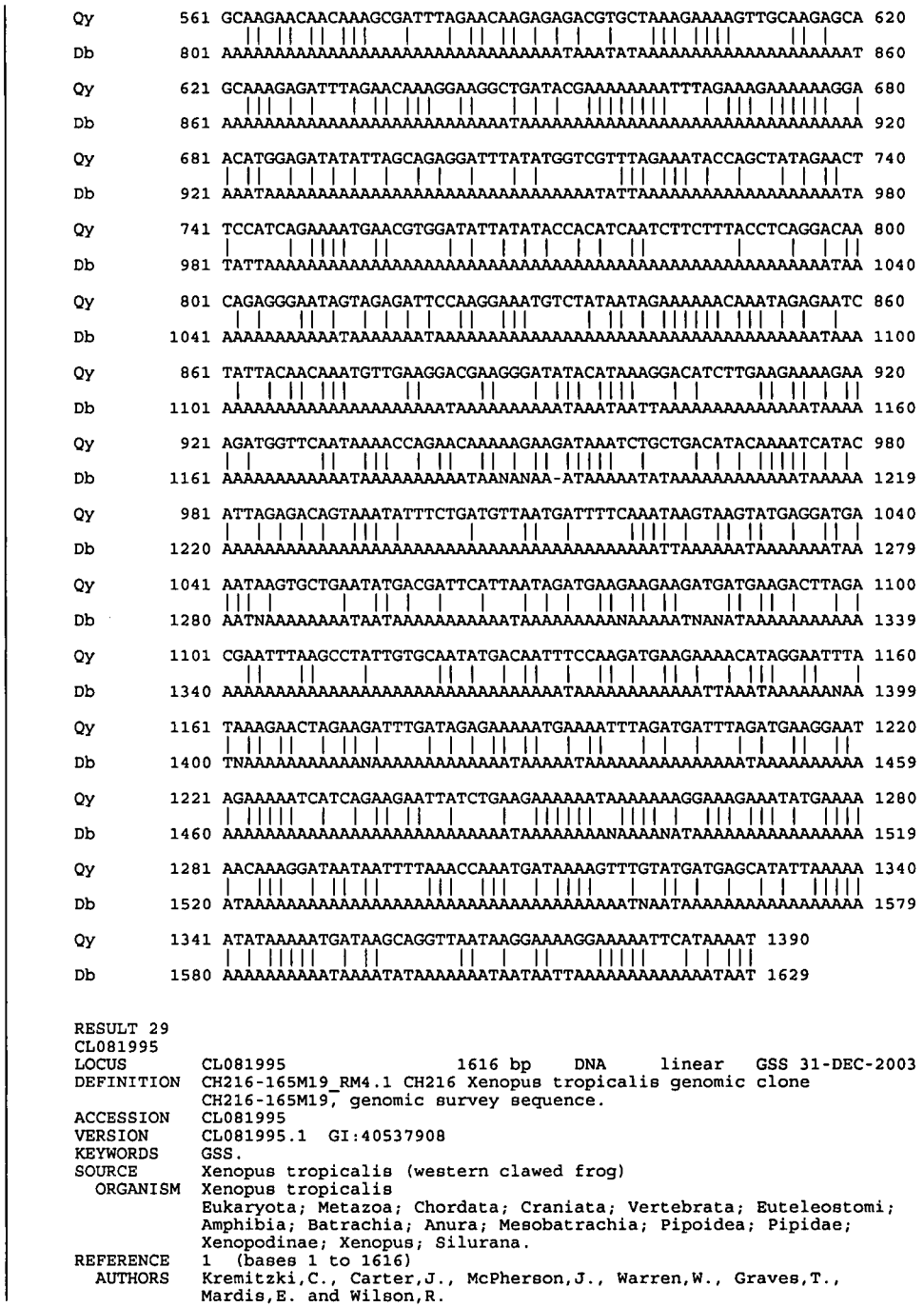


```
RESULT 28
CL082569
LOCUS       CL082569                1632 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-167P5_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-167P5, genomic survey sequence.
ACCESSION   CL082569
VERSION     CL082569.1  GI:40538482
KEYWORDS    GSS.
```

```
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1632)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1114.
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-167P5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN
```

Query Match 15.4%; Score 228.4; DB 10; Length 1632;
Best Local Similarity 47.5%; Pred. No. 7.9e-28;
Matches 651; Conservative 0; Mismatches 718; Indels 1; Gaps 1;





```
RESULT 29
CL081995
LOCUS       CL081995               1616 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-165M19_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-165M19, genomic survey sequence.
ACCESSION   CL081995
VERSION     CL081995.1  GI:40537908
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
```

```
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 1038
            High quality sequence stop: 1108.
FEATURES             Location/Qualifiers
     source          1. .1616
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-165M19"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match           15.4%;  Score 228.2;  DB 10;  Length 1616;
  Best Local Similarity   42.7%;  Pred. No. 8.5e-28;
  Matches  593;  Conservative    0;  Mismatches  795;  Indels    0;  Gaps    0;

Qy          5 AACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 64
              |
Db        199 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 258

Qy         65 GCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAG 124
                                                                 |
Db        259 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANNNANNNNNNNNNNNNNNNN 318

Qy        125 AACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGA 184
              |     |   |      ||| ||||     || || || |||    |   | || || | |
Db        319 ANNNNNANNNAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 378

Qy        185 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 244
                |      ||| ||||       | || || |||    |   | || ||   | |      |
Db        379 AAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAAAANAAAAAAAANNAAAAAAAAAAA 438

Qy        245 AAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGT 304
              || | ||      || || || |||    |    | || || | | |      ||| |||
Db        439 AAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAA 498

Qy        305 TACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAAC 364
              | || |  |   ||    |   | || || | | |      ||| ||||     || ||
Db        499 AAAAAAAAAAAANAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAANAAAAAA 558

Qy        365 AACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCG 424
              || |||    |   | || || | | |      ||| ||||     || || || |||
Db        559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 618

Qy        425 ATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAAC 484
              |   | || || | | |      ||| ||||     || || || |||    |  | ||
Db        619 AAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 678

Qy        485 AAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGAC 544
              || | | |      ||| ||||     |  || || |||    |   | || || | |
Db        679 AAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 738

Qy        545 GTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAG 604
                  |||  |||     || || || |||    |   | || || |   |      |||
Db        739 AAAAAAAAANAAAAAAAAAANAAAAANAAAAAAAAAAAAAAAAAAANNAAAAAAANAAAN 798

Qy        605 AAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATT 664
              ||||    || |  |  ||| | |   | || |||  ||     | |  ||||||||
Db        799 AAAANANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 858

Qy        665 TAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAA 724
              | ||| ||||||  ||     | | | |  | | |  |  | |          | |||
Db        859 AAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAA 918

Qy        725 TACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTT 784
              |  |   | | ||     |  | ||||  ||     | |  | | |  | |  ||
Db        919 NAAAAAAAAAAAAAANNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 978

Qy        785 CTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAA 844
                  |   | || | |   ||     | | |    ||  |||     | || |  ||
Db        979 AAAAAAAAAAAAAAAAAAANAAAAAAANAAAAAAAANAAAAAAAAAAAAAAAAAAAANAA 1038

Qy        845 AAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGAC 904
              ||| ||| | | ||   |  | || |||    ||  |  ||   | | | | |||  |
Db       1039 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1098

Qy        905 ATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTG 964
              |     || ||   ||| |       || ||||  | || ||||| || | |||
Db       1099 AAAAAAAAAAANNAAAAAANAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAA 1158

Qy        965 ACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAA 1024
              | | | ||||  | | |  | | | |  ||| |      |    ||  |     ||| ||
Db       1159 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAA 1218

Qy       1025 GTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAG 1084
                || |  |  |  ||| ||    || | |  |   |  || | |  || || ||
Db       1219 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAA 1278

Qy       1085 ATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAG 1144
              | |  || |   | |  ||   ||      | |  | ||     || | ||   || | ||
Db       1279 AAAAAAAAAAANAAAANAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAANAAAAAAAAAA 1338

Qy       1145 AAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATG 1204
              ||||   |  ||   | ||| ||  | ||       | | | |||||  ||||   | |
Db       1339 AAAAANAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAANAAAA 1398

Qy       1205 ATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAG 1264
              |  | |  ||  || | |||||  |  | || ||  |    || |||||| |||||||
Db       1399 AAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAA 1458

Qy       1265 GAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATG 1324
              ||| ||| |  |||||  ||| | || ||  ||| |||  |   ||      |
Db       1459 AAAAAAAAAAAAAAAAAANAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAA 1518

Qy       1325 ATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCA 1384
              |  | | |  |||||| | |||||  | ||  |      | ||  ||||  |||||   |
Db       1519 AAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAANANAAANAAAANNAAAAAAANA 1578

Qy       1385 TAAAATCA 1392
              || | |
Db       1579 AAANANNA 1586


RESULT 30
CL078686
LOCUS       CL078686                1630 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-152J24_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-152J24, genomic survey sequence.
ACCESSION   CL078686
VERSION     CL078686.1  GI:40534599
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1630)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
```

```
           Washington University School of Medicine
           Email: submissions@watson.wustl.edu
           Insert Length: 175000   Std Error: 0.00
           Seq primer: Sp5 atctgccgtttcgatcct
           Class: BAC ends
           High quality sequence start: 498
           High quality sequence stop: 596.
FEATURES             Location/Qualifiers
     source          1. .1630
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-152J24"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             15.4%;  Score 228;  DB 10;  Length 1630;
  Best Local Similarity   48.8%;  Pred. No. 9.2e-28;
  Matches  677;  Conservative    0;  Mismatches  705;  Indels    6;  Gaps    2;

Qy          2 AAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAAC 61
              ||  ||  |  |||   |   |  || || | | | |    ||  ||||   | || || ||
Db        216 AAAAAAAGGAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAGAAAAAAAAGAAAAAAAAA 275

Qy         62 AAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATT 121
              |||   |   | | ||  | |||      |||  ||||  | || |  || |||   ||
Db        276 AAAAAAAGAGAAATCATAAAAGAAAAAGAAAAAGAAAGAAAAAAAAAAAGAAAA---ATA 332

Qy        122 TAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAG 181
               | || || |  ||      ||| ||||     || || || | |   |   | || ||
Db        333 AAAAAAAAAAAGGAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAA 392

Qy        182 AGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTG 241
              | | |      ||| ||||     || || || |||   |   | || || | | |
Db        393 AAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAA 452

Qy        242 CTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAA 301
              | ||| |||||    || || || |||   |   | || || | | |      ||| |||
Db        453 CAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAA 512

Qy        302 AGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAG 361
              |   | || |   | ||||   |   | || || | | |      ||| ||||     ||
Db        513 AAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAA 572

Qy        362 AACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAA 421
              || || ||||  |   | || || | | ||     ||| ||||     || || || |||
Db        573 AAAAAAAAAGAAAAAAAAAAAAAAAAACACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 632

Qy        422 GCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAG 481
                 |   | || || | | |   | ||| ||||   | || || || |||   |   ||
Db        633 AAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAAAAAAAAAAAA---AAAAAG 689

Qy        482 AACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGA 541
              || || | | |       ||||||||      || || || |||        | || || | |
Db        690 AAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAGAAGAAAAAACAAAAAAAAAAAAAA 749

Qy        542 GACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTA 601
               |    ||| |||||    || || || |||   |  | || || | | |       |
Db        750 AAAAAAAAAAAAAAAAGAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAGA 809

Qy        602 AAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAA 661
              || ||||      || |   |   || | |    |||| |||  ||     || |  |||||||
Db        810 AAAAAAAAAAAAAAAAAAAAAACAAAAAAAAGAGAAAAAAAAAAAAAAGAAAAAAAAAAA 869

Qy        662 ATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAG 721
              |   | ||||||||||   | |   | | | |  || | |   |   | | |   |   |
Db        870 AAAAAAAAAGAAAAAAAACAGAAAAAAACAGAAAAGAAAAAAAAAGAAAAGAAAGAAAAA 929

Qy        722 AAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAAT 781
              |   | |   | | ||     |  | ||||  ||      | |  | | | |  | |  ||
Db        930 ACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 989

Qy        782 CTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAG 841
                    | | || || | |   || |  ||| |    ||  |||     | || |
Db        990 AAAAAGAAAAAAAAGAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAGAAAAA 1049

Qy        842 AAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAG 901
              |||||  |||   | ||   |  | || |||      || ||  ||   | | | | |||
Db       1050 AAAAAGAAAAAGAAAAAAAAAGAAAAAAAAAGAAAAAAAGAAAAAAAAAAAAAGAAAAA 1109

Qy        902 GACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTG 961
               | |      || |||| ||| |         |  |||  ||||  |||||||||  | |
Db       1110 AAAAAAAAAAAAAAAAAAAAAAACAAAAACAAGAAAGAAGAAAGAAAAGAAGAAAGAAAAA 1169

Qy        962 CTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAA 1021
                 | |  | ||  | | |  | | | |  ||| |       |    ||  |      |||
Db       1170 AAAAAAAGAAGAAAGAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAA 1229

Qy       1022 TAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAG 1081
               |   ||    ||  |  ||  ||       || |  |  |    |  | |    |  ||
Db       1230 AAGAAAAAAGAGAAAAAAAAGGAAAGAAAAAAAACAAAAAAAAAAAGAAAAGAAAGAAAA 1289

Qy       1082 AAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATG 1141
               || |  |  ||||   |||  ||   ||     |       || |  | |||    || ||
Db       1290 AAAAAAAGAAAGAACAAGAAAAAGAAAAAGAAAAAAAAGAAAAGAAGAAACAACAGGAGA 1349

Qy       1142 AAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAG 1201
               || |||| | || ||    | | | |||    ||||    | | | ||| |  ||||   |
Db       1350 AAAAAAAAACAGAAAAAGAAAGAAAACACAAAGAAGAAAAAAAAAAAAGAGAAAAAAAAAA 1409

Qy       1202 ATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAA 1261
              | ||   |||  ||  || | |||          || ||  |    |||||||||  || ||
Db       1410 AAGAAAAAGAGAAATAAAAAAAAAGGAAGAGGAAAAAAAAAAAGAAAGAAAAAAAAAGAA 1469

Qy       1262 AAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGT 1321
              ||| ||  ||| |  | |||| ||    | |  ||      ||  ||  ||  |||
Db       1470 AAGAAACAAAAAAAAACAAAAGAACAAAAACAAACGAAGAAAGAAGAGAAGAAAAAAGAA 1529

Qy       1322 ATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAAT 1381
              |  | ||  | |  ||||||   |||||  | ||| |    || ||| ||||  | |||
Db       1530 AAAAAGAAAAAAAGAAAAAAAGAAAAAAAAAAAAAGAAAAGAAAGAAGAAAAAAAAGAAAA 1589

Qy       1382 TCATAAAA 1389
                | ||||
Db       1590 GAAAAAAA 1597


Search completed: December 20, 2005, 15:08:08
Job time : 7887.23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:28:07 ; Search time 372.484 Seconds
(without alignments)
5614.164 Million cell updates/sec

Title: US-09-837-344-45
Perfect score: 2487
Sequence: 1 CAAGAACAACAAAGCGATCT.....AATATTTTATGAAACTATAA 1482

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22916/app_query.fasta_1.3740
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344 @CGN_1_1_1319 @runat_21122005_093215_22916 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2475	99.5	493	2	Q25886_PLAFA	Q25886 plasmodium
2	2403.5	96.6	1909	2	Q25893_PLAFA	Q25893 plasmodium
3	2372.5	95.4	1596	2	Q8IJ44_PLAF7	Q8ij44 plasmodium
4	1448	58.2	280	2	Q25844_PLAFA	Q25844 plasmodium
5	1447	58.2	280	2	Q25843_PLAFA	Q25843 plasmodium
6	1444	58.1	280	2	Q25847_PLAFA	Q25847 plasmodium
7	1444	58.1	280	2	Q25848_PLAFA	Q25848 plasmodium
8	1444	58.1	280	2	Q25852_PLAFA	Q25852 plasmodium
9	1442	58.0	280	2	Q25849_PLAFA	Q25849 plasmodium
10	1442	58.0	280	2	Q25851_PLAFA	Q25851 plasmodium
11	1440	57.9	280	2	Q25850_PLAFA	Q25850 plasmodium
12	1439	57.9	280	2	Q25846_PLAFA	Q25846 plasmodium
13	1434	57.7	280	2	Q25853_PLAFA	Q25853 plasmodium
14	1431	57.5	280	2	Q25854_PLAFA	Q25854 plasmodium
15	1430	57.5	280	2	Q25845_PLAFA	Q25845 plasmodium
16	1429	57.5	280	2	Q25900_PLAFA	Q25900 plasmodium
17	1427	57.4	280	2	Q25855_PLAFA	Q25855 plasmodium
18	1427	57.4	280	2	Q25888_PLAFA	Q25888 plasmodium
19	1425	57.3	280	2	Q27243_PLAFA	Q27243 plasmodium
20	1423	57.2	280	2	Q25889_PLAFA	Q25889 plasmodium
21	1363	54.8	264	2	Q9GTX5_PLAFA	Q9gtx5 plasmodium
22	770.5	31.0	318	2	Q25887_PLAFA	Q25887 plasmodium
23	469.5	18.9	978	2	Q7RAK4_PLAYO	Q7rak4 plasmodium
24	453	18.2	1207	2	Q4UHB4_THEAN	Q4uhb4 theileria a
25	433	17.4	905	2	Q54XN9_DICDI	Q54xn9 dictyosteli
26	429.5	17.3	3427	2	Q6YA79_PLARE	Q6ya79 plasmodium
27	425	17.1	1082	2	Q54JG8_DICDI	Q54jg8 dictyosteli
28	425	17.1	2205	2	Q54C75_DICDI	Q54c75 dictyosteli
29	421	16.9	1061	2	Q54MJ1_DICDI	Q54mj1 dictyosteli
30	419	16.8	762	2	Q50RA3_ENTHI	Q50ra3 entamoeba h
31	419	16.8	1575	2	Q519G4_ENTHI	Q519g4 entamoeba h
32	408	16.4	1738	2	Q51ED7_ENTHI	Q51ed7 entamoeba h
33	403	16.2	1119	2	Q54TT8_DICDI	Q54tt8 dictyosteli
34	400.5	16.1	1108	2	Q9NDI0_BABBI	Q9ndi0 babesia big
35	393.5	15.8	1508	2	Q75JP5_DICDI	Q75jp5 dictyosteli
36	393.5	15.8	2349	2	Q8I455_PLAF7	Q8i455 plasmodium
37	392	15.8	1750	2	Q8X0H2_NEUCR	Q8x0h2 neurospora
38	390	15.7	826	2	Q9VPS3_DROME	Q9vps3 drosophila
39	384.5	15.5	1777	2	Q54FB8_DICDI	Q54fb8 dictyosteli
40	382.5	15.4	1407	1	TRHY_RABIT	P37709 oryctolagus
41	382.5	15.4	1549	1	TRHY_SHEEP	P22793 ovis aries
42	381.5	15.3	1003	2	Q91LX9_HHV8	Q91lx9 human herpe
43	381.5	15.3	1129	2	Q9QR71_HHV8	Q9qr71 human herpe
44	377.5	15.2	1089	2	O40947_HHV8	O40947 human herpe
45	377	15.2	1036	2	Q9DUM3_HHV8	Q9dum3 human herpe

ALIGNMENTS

RESULT 1
Q25886_PLAFA
ID Q25886_PLAFA PRELIMINARY; PRT; 493 AA.
AC Q25886;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage antigen-1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94267224; PubMed=7515922;
RA Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT "Plasmodium falciparum liver stage antigen-1 is well conserved and
RT contains potent B and T cell determinants.";
RL J. Immunol. 153:190-204(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT "The Plasmodium falciparum liver stage antigen LSA-1 is well conserved
RT and harbors major B- and T-cell epitopes.";
RL J. Immunol. 0:0-0(1994).
DR EMBL; Z30320; CAA82975.1; -; Genomic_DNA.
FT NON_TER 1 1
SQ SEQUENCE 493 AA; 59062 MW; FE252CC44F46996E CRC64;

Alignment Scores:

Pred. No.:	2.69e-107	Length:	493
Score:	2475.00	Matches:	492
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.80%	Mismatches:	0
Query Match:	99.52%	Indels:	0
DB:	2	Gaps:	0

```
US-09-837-344-45 (1-1482) x Q25886_PLAFA (1-493)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 20

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAsp 40

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 60

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 80

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGlu 100

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGln 120

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 140

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 160

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 180

Qy        541 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 200

Qy        601 AAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LysGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLys 220

Qy        661 AATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnLeuGluArgLysLysGluHisGlyAspIleLeuAlaGluAspLeuTyrGlyArgLeu 240

Qy        721 GAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GluIleProAlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGln 260

Qy        781 TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATA 840
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        261 SerSerLeuProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIle 280

Qy        841 GAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GluLysThrAsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLys 300

Qy        901 GGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GlyHisLeuGluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSer 320

Qy        961 GCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AlaAspIleGlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGln 340

Qy       1021 ATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        341 IleSerLysTyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGlu 360

Qy       1081 GAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluAspAspGluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAsp 380

Qy       1141 GAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GluGluAsnIleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeu 400

Qy       1201 GATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AspAspLeuAspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLys 420

Qy       1261 AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LysGlyLysLysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeu 440

Qy       1321 TATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 TyrAspGluHisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLys 460

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 PheIleLysSerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAsp 480

Qy       1441 GAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||
Db        481 GluLeuSerGluAspIleThrLysTyrPheMetLysLeu 493

RESULT 2
Q25893_PLAFA
ID   Q25893_PLAFA PRELIMINARY;      PRT;  1909 AA.
AC   Q25893;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Liver stage antigen.
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=NF54;
RX   MEDLINE=92107224; PubMed=1840628; DOI=10.1016/0166-6851(91)90117-O;
RA   Zhu J., Hollingdale M.R.;
RT   "Structure of Plasmodium falciparum liver stage antigen-1.";
RL   Mol. Biochem. Parasitol. 48:223-226(1991).
DR   EMBL; X56203; CAA39663.1; -; Genomic_DNA.
DR   PIR; S24597; A45592.
DR   HSSP; O15813; 1D7M.
SQ   SEQUENCE   1909 AA;  230153 MW;  8B861A524C797217 CRC64;

Alignment Scores:
Pred. No.:              4.16e-104       Length:         1909
Score:                  2403.50         Matches:        486
Percent Similarity:     89.89%          Conservative:   3
Best Local Similarity:  89.34%          Mismatches:     4
Query Match:            96.64%          Indels:         51
DB:                     2               Gaps:           2

US-09-837-344-45 (1-1482) x Q25893_PLAFA (1-1909)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 GlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGln 1385

Qy         61 CAAAGCGATTTAGAACAAGATAGACTT---------------------------------- 87
              |||||||||||||||||||||||||||
Db       1386 GlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGlnGluGlnGlnArgAsp 1405

Qy         88 ------------------GCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 129
```

```
Db      1406 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 1425

Qy       130 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGT 189

Db      1426 GluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArg 1445

Qy       190 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 249

Db      1446 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 1465

Qy       250 AAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAA 309

Db      1466 LysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGln 1485

Qy       310 GAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 369

Db      1486 GluGlnGlnArgAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGln 1505

Qy       370 AGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 429

Db      1506 SerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeu 1525

Qy       430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489

Db      1526 GluGlnGluArgLeuAlaAsnGluLysLeuGlnGluGlnGlnArgAspLeuGluGlnGlu 1545

Qy       490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549

Db      1546 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1565

Qy       550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT------------ 597

Db      1566 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1585

Qy       597 ------------------------------------------------------------ 597

Db      1586 LeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlu 1605

Qy       598 ----------------------------GCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627

Db      1606 GlnGlnArgAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnArg 1625

Qy       628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGA 687

Db      1626 AspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGly 1645

Qy       688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747

Db      1646 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 1665

Qy       748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807

Db      1666 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 1685

Qy       808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867

Db      1686 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 1705

Qy       868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927

Db      1706 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 1725

Qy       928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987

Db      1726 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 1745

Qy       988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047

Db      1746 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 1765

Qy      1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107

Db      1766 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspGluAspLeuAspGluPhe 1785

Qy      1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167

Db      1786 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 1805

Qy      1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227

Db      1806 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 1825

Qy      1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAACAAAG 1287

Db      1826 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 1845

Qy      1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347

Db      1846 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 1865

Qy      1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407

Db      1866 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 1885

Qy      1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467

Db      1886 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 1905

Qy      1468 TTTATGAAACTA 1479

Db      1906 PheMetLysLeu 1909
```

RESULT 3
Q8IJ44_PLAF7
ID Q8IJ44_PLAF7 PRELIMINARY; PRT; 1596 AA.
AC Q8IJ44;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Liver stage antigen, putative.
GN ORFNames=PF10_0356;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35553.1; -; Genomic_DNA.
SQ SEQUENCE 1596 AA; 191691 MW; C0B3A2CC0F5D60AC CRC64;

Alignment Scores:

Pred. No.:	1.18e-102	Length:	1596
Score:	2372.50	Matches:	477
Percent Similarity:	94.71%	Conservative:	6
Best Local Similarity:	93.53%	Mismatches:	10
Query Match:	95.40%	Indels:	17
DB:	2	Gaps:	1

US-09-837-344-45 (1-1482) x Q8IJ44_PLAF7 (1-1596)

```
Qy         1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60

Db      1087 GlnGlyGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGln 1106
```

```
Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              ||||||||||||||||||:::|||||||||||||||||||||||||   ||||||||||||
Db       1107 GlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnGlnSerAsp 1126

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db       1127 LeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnGlnSerAspLeuGluGln 1146

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
              ||||||   ||||||||||||||||||:::|||||||||||||||||||||||||||||
Db       1147 GluArgLeuAlaLysGluLysLeuGlnGluArgGlnSerAspLeuGluGlnGluArgLeu 1166

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1167 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGlu 1186

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
              ||||||||||||||||||||||||||||||||:::|||   ||||||||||||||||||
Db       1187 LysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeuAlaLysGluLysLeuGln 1206

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG---------------- 405
              ||||||||||||||||||||||||||||||||   ||||||||||||
Db       1207 GluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGlyGlnHis 1226

Qy        406 ----------------------------------TTGCAAGAACAACAAAGCGATTTA 429
                                                 |||||||||||||||||||||||||
Db       1227 SerAspLeuGluArgThrLysAlaSerLysGluThrLeuGlnGluGlnGlnSerAspLeu 1246

Qy        430 GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1247 GluGlnGluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGlu 1266

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1267 ArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAla 1286

Qy        550 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 LysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLys 1306

Qy        610 TTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 LeuGlnGluGlnGlnArgAspLeuGluGlnArgLysAlaAspThrLysLysAsnLeuGlu 1326

Qy        670 AGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       1327 ArgLysLysGluHisGlyAspValLeuAlaGluAspLeuTyrGlyArgLeuGluIlePro 1346

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTA 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 AlaIleGluLeuProSerGluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeu 1366

Qy        790 CCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACA 849
              ||||||||||||||||||||||||||||||||||:::|||||||||||||   |||
Db       1367 ProGlnAspAsnArgGlyAsnSerArgAspSerLysGluIleSerIleIleGluAsnThr 1386

Qy        850 AATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTT 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 AsnArgGluSerIleThrThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeu 1406

Qy        910 GAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATA 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1407 GluGluLysLysAspGlySerIleLysProGluGlnLysGluAspLysSerAlaAspIle 1426

Qy        970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 GlnAsnHisThrLeuGluThrValAsnIleSerAspValAsnAspPheGlnIleSerLys 1446

Qy       1030 TATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGAT 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 TyrGluAspGluIleSerAlaGluTyrAspAspSerLeuIleAspGluGluGluAspAsp 1466

Qy       1090 GAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAAC 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1467 GluAspLeuAspGluPheLysProIleValGlnTyrAspAsnPheGlnAspGluGluAsn 1486

Qy       1150 ATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTA 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 IleGlyIleTyrLysGluLeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeu 1506

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1507 AspGluGlyIleGluLysSerSerGluGluLeuSerGluGluLysIleLysLysGlyLys 1526

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1527 LysTyrGluLysThrLysAspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGlu 1546

Qy       1330 CATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAA 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1547 HisIleLysLysTyrLysAsnAspLysGlnValAsnLysGluLysGluLysPheIleLys 1566

Qy       1390 TCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCT 1449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1567 SerLeuPheHisIlePheAspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSer 1586

Qy       1450 GAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||
Db       1587 GluAspIleThrLysTyrPheMetLysLeu 1596
```

RESULT 4
Q25844_PLAFA
ID Q25844_PLAFA PRELIMINARY; PRT; 280 AA.
AC Q25844;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN Name=LSA-1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; L40885; AAB59231.1; -; Genomic_DNA.
FT NON_TER 1 1
SQ SEQUENCE 280 AA; 32940 MW; E9708E3CFFAEA9CF CRC64;

Alignment Scores:

Pred. No.:	1.26e-59	Length:	280
Score:	1448.00	Matches:	279
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.64%	Mismatches:	0
Query Match:	58.22%	Indels:	0
DB:	2	Gaps:	0

US-09-837-344-45 (1-1482) x Q25844_PLAFA (1-280)

```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy      1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 5
Q25843_PLAFA
ID   Q25843_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25843;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40884; AAB59230.1; -; Genomic_DNA.
FT   NON_TER      1      1
SQ   SEQUENCE   280 AA;  32926 MW;  467080F32FAEAD33 CRC64;

Alignment Scores:
Pred. No.:              1.4e-59        Length:        280

Score:                  1447.00        Matches:       278
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  99.29%         Mismatches:    0
Query Match:            58.18%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25843_PLAFA (1-280)

Qy       640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy       700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy       760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy       820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
             |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy       880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy       940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy      1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy      1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy      1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy      1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy      1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy      1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy      1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy      1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 6
Q25847_PLAFA
ID   Q25847_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25847;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
```

```
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40888; AAC41598.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32944 MW;  467085E23BABBC27 CRC64;

Alignment Scores:
Pred. No.:              1.92e-59        Length:         280
Score:                  1444.00         Matches:        277
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.93%          Mismatches:     0
Query Match:            58.06%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x Q25847_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysMetLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 7
Q25848_PLAFA
ID   Q25848_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40889; AAC41599.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32939 MW;  E77080385FAA0D61 CRC64;

Alignment Scores:
Pred. No.:              1.92e-59        Length:         280
Score:                  1444.00         Matches:        278
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  99.29%          Mismatches:     0
Query Match:            58.06%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x Q25848_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LysGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 8
Q25852_PLAFA
ID   Q25852_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25852;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40834; AAB59233.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32927 MW;  E6648F85FBBAA245 CRC64;

Alignment Scores:
Pred. No.:              1.92e-59        Length:         280
Score:                  1444.00         Matches:        277
Percent Similarity:     100.00%         Conservative:   3
Best Local Similarity:  98.93%          Mismatches:     0
Query Match:            58.06%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x Q25852_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                 ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
                 ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                 |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGluLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 9
Q25849_PLAFA
ID   Q25849_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25849;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40890; AAC41600.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32966 MW;  E96D255154DEA9CF CRC64;

Alignment Scores:
Pred. No.:              2.38e-59        Length:         280
Score:                  1442.00         Matches:        278
```

```
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  99.29%         Mismatches:    0
Query Match:            57.98%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25849_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluTyrIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 10
Q25851_PLAFA
ID   Q25851_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40893; AAC41603.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32927 MW;  48706EFDCF40ADD3 CRC64;

Alignment Scores:
Pred. No.:              2.38e-59       Length:        280
Score:                  1442.00        Matches:       277
Percent Similarity:     100.00%        Conservative:  3
Best Local Similarity:  98.93%         Mismatches:    0
Query Match:            57.98%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25851_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAspGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240
```

```
Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 11
Q25850_PLAFA
ID   Q25850_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25850;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40891; AAC41601.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32989 MW;  E96F812CFFAEBED8 CRC64;

Alignment Scores:
Pred. No.:              2.95e-59       Length:         280
Score:                  1440.00        Matches:        278
Percent Similarity:     99.64%         Conservative:   1
Best Local Similarity:  99.29%         Mismatches:     1
Query Match:            57.90%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-45 (1-1482) x Q25850_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspIleLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspTyrAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 12
Q25846_PLAFA
ID   Q25846_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40887; AAC41597.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32882 MW;  46763641C81AFC33 CRC64;

Alignment Scores:
Pred. No.:              3.28e-59       Length:         280
Score:                  1439.00        Matches:        277
Percent Similarity:     99.64%         Conservative:   2
Best Local Similarity:  98.93%         Mismatches:     1
Query Match:            57.86%         Indels:         0
DB:                     2              Gaps:           0

US-09-837-344-45 (1-1482) x Q25846_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60
```

```
Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnAspLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 13
Q25853_PLAFA
ID   Q25853_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25853;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40835; AAB59234.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32938 MW;  46751C45F4DCBD33 CRC64;

Alignment Scores:
Pred. No.:              5.6e-59        Length:        280
Score:                  1434.00        Matches:       276
Percent Similarity:     99.29%         Conservative:  2
Best Local Similarity:  98.57%         Mismatches:    2
Query Match:            57.66%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25853_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 14
Q25854_PLAFA
ID   Q25854_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25854;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40836; AAC41604.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32966 MW;  61851C5AD6E1A211 CRC64;

Alignment Scores:
Pred. No.:              7.71e-59        Length:         280
Score:                  1431.00         Matches:        275
Percent Similarity:     99.29%          Conservative:   3
Best Local Similarity:  98.21%          Mismatches:     2
Query Match:            57.54%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x Q25854_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluArgIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 15
Q25845_PLAFA
ID   Q25845_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25845;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40886; AAB59232.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32908 MW;  FDEE929F34DCA7E9 CRC64;

Alignment Scores:
Pred. No.:              8.58e-59        Length:         280
Score:                  1430.00         Matches:        275
Percent Similarity:     98.93%          Conservative:   2
Best Local Similarity:  98.21%          Mismatches:     3
Query Match:            57.50%          Indels:         0
DB:                     2               Gaps:           0

US-09-837-344-45 (1-1482) x Q25845_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160
```

```
Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 16
Q25900_PLAFA
ID   Q25900_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25900;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40947; AAC41592.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32911 MW;  F8850CABF6136212 CRC64;

Alignment Scores:
Pred. No.:              9.54e-59       Length:        280
Score:                  1429.00        Matches:       275
Percent Similarity:     99.29%         Conservative:  3
Best Local Similarity:  98.21%         Mismatches:    2
Query Match:            57.46%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25900_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValSerAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 17
Q25855_PLAFA
ID   Q25855_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25855;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40837; AAC41605.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32866 MW;  73164A709789DD31 CRC64;

Alignment Scores:
Pred. No.:              1.18e-58       Length:        280
Score:                  1427.00        Matches:       275
Percent Similarity:     98.93%         Conservative:  2
Best Local Similarity:  98.21%         Mismatches:    3
```

```
Query Match:            57.38%      Indels:        0
DB:                     2           Gaps:          0

US-09-837-344-45 (1-1482) x Q25855_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSerAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGlyAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 18
Q25888_PLAFA
ID   Q25888_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25888;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40908; AAB59228.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   280 AA;  32936 MW;  FDF9A29104C5A7E9 CRC64;

Alignment Scores:
Pred. No.:              1.18e-58      Length:        280
Score:                  1427.00       Matches:       274
Percent Similarity:     98.93%        Conservative:  3
Best Local Similarity:  97.86%        Mismatches:    3
Query Match:            57.38%        Indels:        0
DB:                     2             Gaps:          0

US-09-837-344-45 (1-1482) x Q25888_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheArgProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 19
Q27243_PLAFA
ID   Q27243_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q27243;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40910; AAC41582.1; -; Genomic_DNA.
DR   EMBL; L40892; AAC41602.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32881 MW;  105849F296958C78 CRC64;

Alignment Scores:
Pred. No.:              1.46e-58       Length:        280
Score:                  1425.00        Matches:       274
Percent Similarity:     98.93%         Conservative:  3
Best Local Similarity:  97.86%         Mismatches:    3
Query Match:            57.30%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q27243_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
              |||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
              |||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluSerIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 20
Q25889_PLAFA
ID   Q25889_PLAFA PRELIMINARY;      PRT;   280 AA.
AC   Q25889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen 1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=96065765; PubMed=7477115; DOI=10.1016/0166-6851(95)00069-D;
RA   Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA   Hawley W.A., Collins W.E., Lal A.A.;
RT   "Sequence variations in the non-repetitive regions of the liver stage-
RT   specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT   isolates.";
RL   Mol. Biochem. Parasitol. 71:291-294(1995).
DR   EMBL; L40909; AAB59229.1; -; Genomic_DNA.
FT   NON_TER       1       1
SQ   SEQUENCE   280 AA;  32850 MW;  FDEE929F34DF68E9 CRC64;

Alignment Scores:
Pred. No.:              1.81e-58       Length:        280
Score:                  1423.00        Matches:       274
Percent Similarity:     98.57%         Conservative:  2
Best Local Similarity:  97.86%         Mismatches:    4
Query Match:            57.22%         Indels:        0
DB:                     2              Gaps:          0

US-09-837-344-45 (1-1482) x Q25889_PLAFA (1-280)

Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          1 ArgLysAlaAspThrLysLysAsnLeuGluArgLysLysGluHisGlyAspValLeuAla 20

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db         21 GluAspLeuTyrGlyArgLeuGluIleProValIleGluLeuProSerGluAsnGluArg 40

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGlyAsnSerArgAsp 60

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAA 879
```

```
                 ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThrThrAsnValGlu 80

Qy        880 GGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCA 939
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGlySerIleLysPro 100

Qy        940 GAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATT 999
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGluThrValAsnIle 120

Qy       1000 TCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGAC 1059
                 ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        121 SerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleGlyAlaGluTyrAsp 140

Qy       1060 GATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTG 1119
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPheLysProIleVal 160

Qy       1120 CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTG 1179
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGluLeuGluAspLeu 180

Qy       1180 ATAGAGAAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAA 1239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLysSerSerGluGlu 200

Qy       1240 TTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTT 1299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 LeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLysAspAsnAsnPhe 220

Qy       1300 AAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAG 1359
                 ||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        221 LysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLysAsnValLysGln 240

Qy       1360 GTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAAT 1419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePheAspGlyAspAsn 260

Qy       1420 GAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATATTTTATGAAACTA 1479
                 ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        261 GluIleLeuGlnIleValGlyGluLeuSerGluAspIleThrLysTyrPheMetLysLeu 280

RESULT 21
Q9GTX5_PLAFA
ID   Q9GTX5_PLAFA PRELIMINARY;      PRT;   264 AA.
AC   Q9GTX5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Liver stage-specific antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Xu J., Chen S.Y.;
RT   "Cloning and sequence analysis of 3' region of liver stage antigen-1
RT   gene of Plasmodium falciparum isolate FCC1/HN.";
RL   Zhongshan Yi Ke Da Xue Xue Bao 22:410-413(2001).
DR   EMBL; AF246996; AAG12324.1; -; Genomic_DNA.
FT   NON_TER       1      1
SQ   SEQUENCE   264 AA;  31006 MW;  1714D653E8D9D1D7 CRC64;

Alignment Scores:
Pred. No.:              1.11e-55       Length:        264
Score:                  1363.00        Matches:       262
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  99.24%         Mismatches:    0
Query Match:            54.80%         Indels:        0
DB:                     2              Gaps:          0
```

```
US-09-837-344-45 (1-1482) x Q9GTX5_PLAFA (1-264)

Qy        688 GATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCA 747
                 |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AspValLeuAlaGluAspLeuTyrGlyArgLeuGluIleProAlaIleGluLeuProSer 20

Qy        748 GAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGG 807
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAsnGluArgGlyTyrTyrIleProHisGlnSerSerLeuProGlnAspAsnArgGly 40

Qy        808 AATAGTAGAGATTCCAAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACA 867
                 ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 AsnSerArgAspSerLysGluIleSerIleIleGluLysThrAsnArgGluSerIleThr 60

Qy        868 ACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGT 927
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ThrAsnValGluGlyArgArgAspIleHisLysGlyHisLeuGluGluLysLysAspGly 80

Qy        928 TCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 SerIleLysProGluGlnLysGluAspLysSerAlaAspIleGlnAsnHisThrLeuGlu 100

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ThrValAsnIleSerAspValAsnAspPheGlnIleSerLysTyrGluAspGluIleSer 120

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTT 1107
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AlaGluTyrAspAspSerLeuIleAspGluGluGluAspAspGluAspLeuAspGluPhe 140

Qy       1108 AAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAA 1167
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysProIleValGlnTyrAspAsnPheGlnAspGluGluAsnIleGlyIleTyrLysGlu 160

Qy       1168 CTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAA 1227
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuGluAspLeuIleGluLysAsnGluAsnLeuAspAspLeuAspGluGlyIleGluLys 180

Qy       1228 TCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG 1287
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SerSerGluGluLeuSerGluGluLysIleLysLysGlyLysLysTyrGluLysThrLys 200

Qy       1288 GATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAA 1347
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspAsnAsnPheLysProAsnAspLysSerLeuTyrAspGluHisIleLysLysTyrLys 220

Qy       1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAATTCATAAAATCATTGTTTCATATATTT 1407
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnAspLysGlnValAsnLysGluLysGluLysPheIleLysSerLeuPheHisIlePhe 240

Qy       1408 GACGGAGACAATGAAATTTTACAGATCGTGGATGAGTTATCTGAAGATATAACTAAATAT 1467
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AspGlyAspAsnGluIleLeuGlnIleValAspGluLeuSerGluAspIleThrLysTyr 260

Qy       1468 TTTATGAAACTA 1479
                 ||||||||||||
Db        261 PheMetLysLeu 264

RESULT 22
Q25887_PLAFA
ID   Q25887_PLAFA PRELIMINARY;      PRT;   318 AA.
AC   Q25887;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Liver stage antigen-1 (Fragment).
GN   Name=LSA-1;
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=94267224; PubMed=7515922;
```

```
RA   Fidock D.A., Gras-Masse H., Lepers J., Brahimi K., Benmohamed L.,
RA   Mellouk S., Guerin-Marchand C., Londono A., Raharimalala L.,
RA   Meis J.F., Langsley G., Roussilhon C., Tartar A., Druilhe P.;
RT   "Plasmodium falciparum liver stage antigen-1 is well conserved and
RT   contains potent B and T cell determinants.";
RL   J. Immunol. 153:190-204(1994).
DR   EMBL; Z30319; CAA82974.1; -; Genomic_DNA.
DR   HSSP; O15813; 1D7M.
FT   NON_TER     318     318
SQ   SEQUENCE   318 AA;  38098 MW;  989F32A70605DFB2 CRC64;

Alignment Scores:
Pred. No.:              3.15e-28        Length:         318
Score:                  770.50          Matches:        163
Percent Similarity:     92.59%          Conservative:   12
Best Local Similarity:  86.24%          Mismatches:     13
Query Match:            30.98%          Indels:         1
DB:                     2               Gaps:           1

US-09-837-344-45 (1-1482) x Q25887_PLAFA (1-318)

Qy         16 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA 75
Db        131 AspAspAspLysLysLysTyrIleLysGlyGlnAspGluAsnArgGlnGluAspLeuGlu 150

Qy         76 CAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGA 135
Db        151 ---GluLysAlaAlaLysGluLysLeuGlnGlyGlnGlnSerAspSerGluGlnGluArg 169

Qy        136 CTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAA 195
Db        170 ArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLys 189

Qy        196 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 255
Db        190 GluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeu 209

Qy        256 CAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 315
Db        210 GlnGluGlnGlnSerAspLeuGluGlnGluArgLeuAlaLysGluLysLeuGlnGluGln 229

Qy        316 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 375
Db        230 GlnSerAspLeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAsp 249

Qy        376 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 435
Db        250 LeuGluGlnGluArgArgAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGln 269

Qy        436 GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 495
Db        270 GluArgLeuAlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnAspArgLeu 289

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
Db        290 AlaLysGluLysLeuGlnGluGlnGlnSerAspLeuGluGlnGluArgArgAlaLysGlu 309

Qy        556 AAGTTGCAAGAACAACAAAGCGATTTA 582
Db        310 ArgLeuGlnGluGlnGlnSerAspLeu 318

RESULT 23
Q7RAK4_PLAYO
ID   Q7RAK4_PLAYO PRELIMINARY;      PRT;   978 AA.
AC   Q7RAK4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Glutamine-asparagine rich protein.
GN   Name=PY06496;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=17XNL;
RX   MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01002201; EAA18721.1; -; Genomic_DNA.
SQ   SEQUENCE   978 AA;  114234 MW;  B72DC75FDDEB7A27 CRC64;

Alignment Scores:
Pred. No.:              2.24e-14        Length:         978
Score:                  469.50          Matches:        152
Percent Similarity:     57.14%          Conservative:   124
Best Local Similarity:  31.47%          Mismatches:     134
Query Match:            18.88%          Indels:         73
DB:                     2               Gaps:           23

US-09-837-344-45 (1-1482) x Q7RAK4_PLAYO (1-978)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        318 GlnSerGluGlnGluLysLeuGluGluLysArgLeuGluGluGluLysLeuLysGlnGlu 337

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
Db        338 Lys------LeuLysGlnGluLysLeuAlaGlnGluLysLeuAlaGlnGluArg------ 353

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
Db        354 LeuAlaGlnGluLysLeuAlaGlnGluLysLeuLysGlnGluGln------LeuGluArg 371

Qy        181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGT 240
Db        372 GluLysLeuAlaGlnGluLysLeuLysGlnGluLys------LeuLysGlnGluGlnLeu 389

Qy        241 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAA 300
Db        390 AlaGlnGluLys--------------------LeuLysGlnGluLysLeuLysGlnGlu 402

Qy        301 AAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA 360
Db        403 LysLeuLysGlnGluGln------LeuGluArgGluArgLeuAlaGlnGluLysLeuLys 420

Qy        361 GAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 420
Db        421 GlnGluGln------LeuGluGlnGluArgLeuAlaGlnGluLysLeuLysGlnGluGln 438

Qy        421 AGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 480
Db        439 ------LeuGluArgGluArgLeuAlaGlnGluLysLeuLysGlnGluGln------Leu 454

Qy        481 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA---------CAAAGCGATTTA 531
Db        455 GluGlnGluArgLeuAlaGlnGluLysLeuLysGlnGluArgLeuAlaGlnGluLysLeu 474

Qy        532 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTA 582
Db        475 GluArgGluLysLeuGluArgGluLysLeuGluArgGluLysIleGluArgGluLysIle 494
```

```
Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
Db        495 GluArgGluLysIleGluArgGluLysIleGluArgGluGln------LeuGluLysLys 512

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
Db        513 LysLeuGlu---GluLysArgLeuGluGluLysArgLeuGluGluLysArgLeuGluGlu 531

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
Db        532 Lys---------ArgLeuGluGluLysArgIleGluGluGluLysArgIleLysGluGlu 548

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGA---GGGAATAGTAGAGAT 819
Db        549 GlnArgIleLysGluGluGlnArgIleGluGluGluLysArgIleLysGluGluLysArg 568

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAAACAAATAGA------GAATCTATTACAACAAAT 873
Db        569 IleLysGluGluGlnArgIleGluGluGluLysArgIleLysGluGluLysArgIleLys 588

Qy        874 GTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db        589 GluGluLysArgIleLysGluGluLysArgIleLysGluGluLysArg---------Ile 605

Qy        934 AAACCAGAACAAAAA------GAAGATAAATCTGCTGACATACAAAATCATACATTAGAG 987
Db        606 GluGluGluLysLysLeuGluLysGluLysLysProGluLeuGluAsnLysGluArgGly 625

Qy        988 ACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGT 1047
Db        626 ThrIleHis-----------------GlnValLeuSerLysIleLysGluLysLeuGln 639

Qy       1048 GCTGAATATGACGATTCATTAATAGATGAAGAA------GAAGATGATGAAGACTTAGAC 1101
Db        640 ProGluSerGlyLysValValGluGlyGluLysLysValGluAlaGlyLysValValAsp 659

Qy       1102 GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTAT 1161
Db        660 AlaGlyLysValVal----------------------------AspAlaGlyLysVal 669

Qy       1162 AAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGAT----------------- 1203
Db        670 ValGluAlaGlyLysValValAspAlaGlyLysValValAspAlaGlyLysValValGlu 689

Qy       1204 ---------------GATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
Db        690 AlaGlyLysValValGluAlaGlyLysValValGluAlaGlyLysValValGluGlyGlu 709

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
Db        710 LysLysValGluGluGlyLysLysValGluProGluLysLysGluAsn---GluThrLeu 728

Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
Db        729 AspLysAlaLeuSerLysThrGlnGlnGlnGlnLeuGluGlyGluGlnLysValGluGly 748

Qy       1369 GAAAAGGAA 1377
Db        749 GluGlnLys 751
```

RESULT 24
Q4UHB4_THEAN
ID   Q4UHB4_THEAN PRELIMINARY;   PRT; 1207 AA.
AC   Q4UHB4;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=TA20215;
OS   Theileria annulata.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC   Theileria.
OX   NCBI_TaxID=5874;

RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Ankara isolate clone C9;
RA   Pain A., Renauld H., Murphy L., Harris D.A., Quail M.A., Berriman M.,
RA   Hall N., Barrell B.G.;
RT   "The chromosome 1 sequence of Theileria annulata.";
RL   Submitted (APR-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR940347; CAI73525.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1207 AA; 144117 MW; B7D707AF2E1DB7A2 CRC64;

Alignment Scores:

Pred. No.:	1.24e-13	Length:	1207
Score:	453.00	Matches:	151
Percent Similarity:	51.93%	Conservative:	145
Best Local Similarity:	26.49%	Mismatches:	164
Query Match:	18.21%	Indels:	110
DB:	2	Gaps:	25

US-09-837-344-45 (1-1482) x Q4UHB4_THEAN (1-1207)

```
Qy          1 CAAGAAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        272 GlnArgArgGlnArgArgGluGluArgGluArgLysLysGlnGluArgLeuArgLysGln 291

Qy         61 CAAAGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGC 117
Db        292 GluGluGluArgLeuGluGlnGluArgLeuGluLysGluArgLeuAlaGluGlnGluArg 311

Qy        118 ---GATTTAGAACAA---------------------------------------------- 129
Db        312 LeuAspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGlu 331

Qy        130 ------GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAAGAG 183
Db        332 GluThrGluArgLeuLysLysGluArgLeuGluGlnGluArg------LeuGluLysGlu 349

Qy        184 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT-----------------TTA 225
Db        350 ArgLeuGluLysGluArgLeuArgLysLeuGluGluGluArgLeuGluLysGluArgLeu 369

Qy        226 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAT 285
Db        370 GluGlnGluArgLeuGluGlnAlaGlyIleGluGluGluGlnLysArgLeuGluGlnGlu 389

Qy        286 AGACTTGCT----------------------AAAGAAAAGTTACAAGAGCAGCAAAGC 321
Db        390 ArgLeuAlaIleGluGluGlnGluLysLeuGluLysGluArgIleArgLysGluGluGlu 409

Qy        322 GATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---GATTTA 378
Db        410 ArgLeuGluGlnGluArgLeuGluLysGluArgLeuAlaGluGlnGluArgLeuAspIle 429

Qy        379 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 438
Db        430 GluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGluGluThrGlu 449

Qy        439 AGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 498
Db        450 ArgLeuLysLysGluArgLeuGluGlnGluArg------LeuGluLysGluArgLeuGlu 467

Qy        499 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 558
Db        468 LysGluArgLeuGluGlnGlnArgGlnGluGlnGluArgLeuArgLysLeuGluGluArg 487

Qy        559 TTGCAAGAACAACAAAGCGATTTA---GAACAAGAGAGACGTGCTAAAGAAAAGTTG--- 612
Db        488 LeuGluGlnGluArgLeuAlaIleGluGluGlnGluArgLeuGluLysGluArgIleGlu 507

Qy        613 CAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA 672
Db        508 GlnGluArgIleArgLysLeuGluGluGlnArgLeuGluLysGluArgLeuAlaGlu--- 526
```

```
Qy        673 AAAAAGGAACATGGAGATATA---------------TTAGCAGAGGATTTATATGGTCGT 717
Db        527 ---LysGluArgLeuAspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArg 545

Qy        718 TTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACAT 777
Db        546 Leu-----------------AlaArgGluGluThrGluArg------------LeuLys 555

Qy        778 CAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
Db        556 LysGluArgLeuGluGlnGluArgLeuLysGluArgLeuGluLysGluArgLeuGlu 575

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        576 GlnGlnArgGlnGluGlnGluArgLeu---ArgLysLeuGluGluArg------LeuGlu 592

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db        593 LysGluArgIleHisGluGluGlnGluArgLeuGluLysGluArgIleGluGlnGluArg 612

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
Db        613 IleArgLysLeuGluGluGlnArgLeuGluLysGluArgLeuAlaGluLysGluArgLeu 632

Qy       1018 CAAATA--------------------AGTAAGTATGAGGATGAAATAAGTGCTGAATAT 1056
Db        633 AspIleGluGluLysIleArgPheAlaGlnGluValGlnLysArgLeuAlaArgGluGlu 652

Qy       1057 GACGATTCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATT 1116
Db        653 ThrGluArgLeuLysLysGluArgLeuGluGlnGluArgLeuGlu------LysGluArg 670

Qy       1117 GTGCAATATGACAATTTC------------CAAGATGAAGAAAACATAGGAATTTATAAA 1164
Db        671 LeuGluLysGluArgLeuGluGlnGlnArgGlnGluGlnGluArgLeu---------Arg 687

Qy       1165 GAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGAATAGAA 1224
Db        688 LysLeuGluGluArgLeuGluLys---GluArgIleHisGluGluGlnGluArgLeuGlu 706

Qy       1225 AAATCATCAGAAGAA-----------------TTATCTGAAGAAAAAATAAAAAAAGGA 1266
Db        707 LysGluArgIleGluGlnGluArgIleArgLysLeuGluGluGlnArgLeuGluLysGlu 726

Qy       1267 AAGAAATATGAAAAA---------------------------------------------- 1281
Db        727 ArgLeuAlaGluLysGluArgLeuAspIleGluGluLysIleArgPheAlaGlnGluVal 746

Qy       1282 ------------ACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
Db        747 GlnLysArgLeuAlaArgGluGluThrGluArgLeuLysLysGluArgLeuGluGlnGlu 766

Qy       1330 CATATTAAAAAATATAAAAATGATAAGCAG---GTTAATAAGGAAAAGGAAAAATTCATA 1386
Db        767 ArgLeuGluLysGluArgLeuGluLysGluArgLeuArgLysGlnGluGluGluArgLeu 786

Qy       1387 AAA--------------------TCATTGTTTCATATATTTGACGGAGACAATGAAATT 1425
Db        787 GluLysGluArgLeuSerGlyIleAsnValGlyAsnLysArgSerGlyAspTyrAspIle 806

Qy       1426 TTACAGATCGTGGATGAGTTATCTGAAGAT 1455
Db        807 ProAlaPheValProSerIleSerAlaAsp 816
```

RESULT 25
Q54XN9_DICDI
ID   Q54XN9_DICDI PRELIMINARY;   PRT;   905 AA.
AC   Q54XN9;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0206221;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
DR   EMBL; AAFI01000053; EAL68014.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   905 AA; 105156 MW; AEE07236412B83CC CRC64;

Alignment Scores:

Pred. No.:	1.12e-12	Length:	905
Score:	433.00	Matches:	146
Percent Similarity:	51.77%	Conservative:	103
Best Local Similarity:	30.35%	Mismatches:	150
Query Match:	17.41%	Indels:	82
DB:	2	Gaps:	21

US-09-837-344-45 (1-1482) x Q54XN9_DICDI (1-905)

```
Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
Db        188 LysGluAsnLysGluArgLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGlu 207

Qy         61 CAAAGCGAT---------TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
Db        208 ArgLeuGluLysGluLysLeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGlu 227

Qy        112 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 171
Db        228 Arg------LeuGluLysGluArgLeuGluAlaGluArgPheGluLysGluArg------ 243

Qy        172 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------- 222
Db        244 LeuGluSerGluArgLeuGluLysGluArgLeuGluAlaGluArgLeuGluLysGluLys 263

Qy        223 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 282
Db        264 LeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGluArg------LeuGluLys 281

Qy        283 GATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGT 342
Db        282 GluArgLeuGluAlaGluArgLeuGluLysGluArg------LeuGluAlaGluArgLeu 299

Qy        343 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGT 393
Db        300 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluLysLeuGluAlaGluArgLeu 319

Qy        394 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACTT 444
```

```
Db        320 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGluArgLeu 339

Qy        445 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---------TTAGAACAAGAGAGACGT 495
              |||||||||||||:::   ::::::   :::         ||||||   ||||||
Db        340 GluLysGluLysLeuGluAlaGluArgLeuGluLysGluArgLeuGluAlaGluArgLeu 359

Qy        496 GCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAA 555
              |||||||||||||:::   ::::::      ||||||:::||||||         |||
Db        360 GluLysGluLysLeuGluAlaGluArg------LeuGluLysGluArgIleGluAsnGlu 377

Qy        556 AAGTTGCAAGAACAACAAAGCGAT-----------------TTAGAACAAGAGAGACGT 597
              |||:::   ::::::   :::                  :::|||   ||||||
Db        378 IleLeuGluAlaGluArgLeuGluArgLeuGluMetGluLysIleGluIleGluArgLeu 397

Qy        598 GCTAAAGAAAAGTTGCAAGAGCAGCAA---AGAGATTTAGAACAAAGGAAGGCTGATACG 654
              |||::::::||||||||||:::|||   ::::::   |||   ::::::   :::
Db        398 GluLysGlnArgLeuGlnGluGluGlnIleLysGluLysGluMetLysArgLeuGluLys 417

Qy        655 AAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGT 714
              :::|||   ||||||   :::||||||
Db        418 GluLysArgLeuGluHisGlnLysGlu------------------------------- 426

Qy        715 CGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCA 774
              ||||||              |||      |||      |||
Db        427 ---LeuGlu------------ArgLeuAspArgGluArgIleArgAla------------ 437

Qy        775 CATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCT 834
                  |||   ::::::                  |||      ||||||
Db        438 ------GluArgLeuGluLysGluArgLeuGluAlaGluArgLeuGluLysGluArgLeu 455

Qy        835 ATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATA 894
              :::|||:::|||       :::|||                        ::::::|||
Db        456 ValIleLysLysLeuLysLysGlu-------------------------LysGluIle 466

Qy        895 CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGAT 954
                  ||||||   ::::::      |||:::|||   :::||||||::::::
Db        467 ---------GlnLeuGluAsnGluArgLeuGlnSerLeuLysGlnLysGlnLysGlnGlu 483

Qy        955 AAATCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGAT 1014
              :::      :::   :::             |||               |||   :::|||
Db        484 GluGluGluGluGluGluGluGluGluGluGlyGluGluGluAspAspAspAsp 503

Qy       1015 TTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAATATGACGATTCATTAATAGAT 1074
                  :::   |||:::|||      :::      ::::::         :::
Db        504 GluIleGluGluGluGluGluGluGlnLysSerAsnValGluGluGlyGluGluGlu 523

Qy       1075 GAAGAAGAAGATGATGAAGACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTC 1134
              |||||||||:::::|||:::   |||:::         |||   |||   :::
Db        524 GluGluGluGluGluGluAspAspAsp------GluIleLysGlnGluGluGluGlu 541

Qy       1135 CAAGATGAAGAAAACATAGGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAA 1194
              ::::::||||||                        ||||||   |||         :::
Db        542 GluGluGluGluGlu--------------------GluAspSerIleIleThrProLys 554

Qy       1195 AATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAA--- 1251
              |||       :::   |||   :::                 |||||||||      :::
Db        555 AsnSerSerIleIleLysGluGluValThrSerAspIleGluGluLeuGluThrAspThr 574

Qy       1252 ------AAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCA 1305
                    :::                      |||   |||   :::   :::
Db        575 SerPheGlnTyrSerSerAlaAsnSerSerMetMetThrAlaAspGlnAspGluGluGlu 594

Qy       1306 AATGATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAAT 1365
              ||||||            :::      ||||||   ::::::   |||   |||
Db        595 GluAspLys------------AspAspLysLysLysAsnGluSerSerLysGlyValThr 610

Qy       1366 AAG 1368
              |||
Db        611 Lys 611

RESULT 26
Q6YA79_PLARE
ID   Q6YA79_PLARE PRELIMINARY;      PRT;  3427 AA.
AC   Q6YA79;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Reticulocyte binding-like protein 2b.
GN   Name=RBL2b;
OS   Plasmodium reichenowi.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5854;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=14698440; DOI=10.1016/j.molbiopara.2003.10.017;
RA   Rayner J.C., Huber C.S., Galinski M.R., Barnwell J.W.;
RT   "Rapid evolution of an erythrocyte invasion gene family: the
RT   Plasmodium reichenowi Reticulocyte Binding Like (RBL) genes.";
RL   Mol. Biochem. Parasitol. 133:287-296(2004).
DR   EMBL; AY166677; AAO38039.1; -; Genomic_DNA.
SQ   SEQUENCE   3427 AA;  404039 MW;  F7C66ED1296686C1 CRC64;

Alignment Scores:
Pred. No.:              1.22e-12        Length:         3427
Score:                  429.50          Matches:        140
Percent Similarity:     50.18%          Conservative:   136
Best Local Similarity:  25.45%          Mismatches:     171
Query Match:            17.27%          Indels:         103
DB:                     2               Gaps:           24

US-09-837-344-45 (1-1482) x Q6YA79_PLARE (1-3427)

Qy           1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
               ||||||:::            |||:::||||||:::   :::|||:::::::::::::::
Db        2672 GlnGluLysLeuLysAlaGlnGluGluGluArgLysGluArgGluArgIleGluLysGlu 2691

Qy          61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
               :::          |||::::::||||||   :::||||||   :::|||::::::
Db        2692 LysGln------GluLysGluArgLeuGluArgGluLysGlnGluGluGluArgLys--- 2708

Qy         121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT--------- 171
               |||:::|||||||||   :::||||||   :::|||::::::   :::
Db        2709 ---GluArgGluArgLeuGluArgGluLysGlnGluGluGluArgLysGluLysGluArg 2727

Qy         172 CTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 231
               :::|||:::|||::::::   ||||||:::|||:::   ::::::         |||:::
Db        2728 IleGluLysGluLysGlnGluLysGluArgLeuGluArgGluLysGln------GluGlu 2745

Qy         232 GAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAACAAAGCGATTTAGAACAAGATAGA 288
               ||||||:::   :::|||:::|||:::   |||:::|||            :::::::::
Db        2746 GluArgLysGluArgGluArgLeuGluArgGluLysGlnGlnGlnGlnLysGluGlu 2765

Qy         289 CTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAA 348
                  |||   :::   ||||||:::|||:::   |||::::::|||         |||
Db        2766 GluLeuLysArgGlnGluGlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLys 2785

Qy         349 GAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 408
                  :::   ||||||:::|||:::   |||::::::|||         |||   :::
Db        2786 ArgGlnGluGlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGlu 2805

Qy         409 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAA 468
               ||||||:::|||:::   |||::::::|||         |||   :::   ||||||:::
Db        2806 GlnGluLysGlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLys 2825

Qy         469 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
               |||:::   |||::::::|||         |||   :::   ||||||:::|||:::
Db        2826 GlnAlaGlnLeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLysGlnAlaGln 2845

Qy         529 TTAGAACAAGAG---------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 579
               |||::::::|||         :::|||   :::::::::   ||||||||||||
Db        2846 LeuGlnLysGluGluGluLeuLysArgGlnGluGlnGluLysGlnGluGlnGlnGlnLys 2865

Qy         580 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAA 639
                  |||   :::|||:::   :::||||||   :::::::::||||||::::::::::::
Db        2866 GluGluLeuLysArgGlnGluGlnGluLysGlnGluGlnGluGlnArgAsnIleGlnGlu 2885
```

```
Qy        640 AGGAAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCA 699
                                  ||||||   :::|||:::      :::|||:::
Db       2886 ------------------------LeuGluGluGlnLysLys---ProGluIleIleAsn 2896

Qy        700 GAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGT 759
              |||   |||             :::|||             |||          ::::::|||
Db       2897 GluAlaLeuValLysGlyAspLysIle------------LeuGluGlySerAspGlnArg 2912

Qy        760 GGATATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGAT 819
                  :::      :::   ::::::      ||||||      ||||||
Db       2913 AsnMetGluLeu---SerLysProAsnIleSerMetAspAsnThrAspAsnSer---Pro 2930

Qy        820 TCCAAGGAAATGTCTATAATAGAAAAA---------ACAAATAGAGAATCTATTACAACA 870
              |||           |||||||||              |||   |||||||||   |||
Db       2931 SerSerAsnSerGluIleIleGluSerAspAspIleAspAsnSerGluSerIleHisThr 2950

Qy        871 ------------AATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAAAAG 918
                     ::::::|||      :::                 :::|||:::
Db       2951 SerHisMetSerAspIleGluSerThrGlnThrAlaAspIleAsnAsnLeuGlnSerThr 2970

Qy        919 AAAGATGGTTCAATAAAACCAGAACAA----------------------------AAA 948
              :::         |||         |||                             :::
Db       2971 GlnThrValAspIleAsnAsnProGlnSerThrGlnThrAlaAspIleAsnAsnLeuGln 2990

Qy        949 GAAGATAAATCTGCTGACATACAAAAT--------------------------CATACA 981
                   ::::::|||||||||   |||                          :::
Db       2991 SerThrGlnThrAlaAspIleAsnAsnLeuGlnSerThrGlnThrAlaAspIleAsnAsn 3010

Qy        982 TTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGAT--- 1038
              |||::::::         ||||||:::|||:::   |||   ::::::      |||
Db       3011 LeuGlnSerThrGlnThrSerAspIleAsnAsnLeuGlnSerThrGlnThrAlaAspIle 3030

Qy       1039 ------------------------------------------------GAAATAAGT 1047
                                                             |||||||||
Db       3031 AsnAsnLeuGlnSerThrGlnThrSerHisArgSerAsnThrHisGlyGlnGluIleSer 3050

Qy       1048 GCTGAATATGACGATTCATTA--------------ATAGATGAAGAAGAAGATGATGAA 1092
                    ||||||   :::               |||   :::|||:::
Db       3051 AspIleValAspAspGlnIleThrHisProSerAsnIleGlyLysGluLysIleThrHis 3070

Qy       1093 GACTTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATA 1152
              :::   ||||||               :::   :::|||      |||
Db       3071 Asn---AspGluIleSerIleThrSerGluArgAsnAsnIleSerAsp------------ 3085

Qy       1153 GGAATTTATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGAT 1212
                        :::   |||      |||   :::         ::::::   |||
Db       3086 --------------ValAsnAspHisSerGluSerSerAsnIlePheGluAsnGlyAsp 3100

Qy       1213 GAAGGAATAGAAAAATCATCAGAAGAATTATCT-----------GAAGAAAAAATAAAA 1260
                   |||      |||:::   :::   |||           |||      |||
Db       3101 SerThrIleAsnThrSerThrArgAspThrSerSerThrHisAspGluAlaHisIleSer 3120

Qy       1261 ---AAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGT 1317
                  |||      |||::::::         :::|||   |||         ::::::
Db       3121 ProSerGlyAsnAlaTyrAspArgValValSerAspAsnLysLysSerMetGlyGluAsn 3140

Qy       1318 TTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAGGAA 1377
              :::   |||            |||   |||   |||:::   :::      :::::::::
Db       3141 IleLysAsp-----------LysLeuLysIleAspGluSerIleThrThrAspGluGln 3156

Qy       1378 AAATTCATAAAATCATTGTTTCATATATTTGACGGAGACAATGAAATTTTACAGATCGTG 1437
                   |||:::                  |||   ::::::   |||::::::|||
Db       3157 ------IleArg--------------LeuAspAspAsnSerAsnIleIleArgIleAsp 3169

Qy       1438 GATGAGTTATCTGAAGATATAACTAAATAT 1467
                        |||::::::   :::
Db       3170 SerThrAspGlnArgAspValSerSerHis 3179

RESULT 27
Q54JG8_DICDI
ID   Q54JG8_DICDI PRELIMINARY;      PRT;  1082 AA.
AC   Q54JG8;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0187762;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000164; EAL63379.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1082 AA;  127407 MW;  9585384E168072C4 CRC64;

Alignment Scores:
Pred. No.:              2.53e-12        Length:         1082
Score:                  425.00          Matches:        141
Percent Similarity:     50.38%          Conservative:   123
Best Local Similarity:  26.91%          Mismatches:     158
Query Match:            17.09%          Indels:         102
DB:                     2               Gaps:           22

US-09-837-344-45 (1-1482) x Q54JG8_DICDI (1-1082)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG---------CAA 54
              |||:::|||      ||||||:::|||::::::   :::|||||||           |||
Db        583 GluArgGlnArgLysLeuGluGluGluLysLysGluGlnGluLysLysLysLeuGluGln 602

Qy         55 GAACAACAAAGCGATTTAGAACAAGAT---AGACTTGCTAAAGAAAAGTTACAAGAGCAG 111
              |||:::::::::   ||||||   :::   ||||||   :::||||||   |||::::::
Db        603 GluArgGluAlaLysLeuGluAsnGluArgArgLeuArgGluGluLysLysGlnLysGlu 622

Qy        112 CAAAGCGATTTAGAACAA---GAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 168
              :::::::|||   |||:::   |||||||||       |||:::|||   :::::::::
Db        623 ArgAlaAspArgGluGluLysGluArgLeuThrValGluArgLeuAlaLysGluArgGlu 642

Qy        169 GATCTAGAACAAGAGAGACGTGCTAAAGAAAAG-----------------TTGCAAGAA 210
                 ::::::|||||||||      |||||||||                 ::::::
Db        643 LeuValGlnGlnGluArgGluLeuLysGluLysGlnLeuLeuAspLysIleLysLysAsp 662

Qy        211 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 270
              :::|||       ||||||:::|||:::       ||||||:::|||:::   ::::::
Db        663 LysGlnGluLysLeuGluLysGluLysLeuGluLysGluArgLeuGluAsnGluLys--- 681

Qy        271 GATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT------ 324
                 ||||||   :::||||||   ||||||||||||:::::::::::   :::
Db        682 ---LeuGluAsnGluArgLeuGlyLysGluLysLeuGluLysGluLysLeuGluAsnGlu 700

Qy        325 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------ 375
                 ||||||   ||||||         |||:::|||:::   ::::::   :::
```

```
Db        701 ArgLeuGluAsnGluArgLeuGluAsnGluArgLeuGluAsnGluArgLeuGluAsnGlu 720

Qy        376 ---TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT------ 426
                  ||||||   ||||||      ||||||:::|||:::::::::::::   :::
Db        721 ArgLeuGluAsnGluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluAsnGly 740

Qy        427 ------TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTG---------------------- 459
                    ||||||:::|||:::|||   ||||||||||||
Db        741 GluLysLeuGluLysGluLysLeuGluLysGluLysLeuGluLysGluGluLeuGluAsn 760

Qy        460 ----------------------------CAAGAACAACAAAGCGATTTAGAACAAGAG 489
                                           |||:::||||||       ::::::|||
Db        761 LysArgAsnGluGluGluLeuLeuLeuLysGlnGlnGlnGlnGlnLysGluLysGluGlu 780

Qy        490 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCT 549
                  :::      |||||||||   ::::::::::::   :::   |||:::|||   :::
Db        781 LysGluGluLysGluLysGluGluLysGluLysGluGluLysGluLysGluGluLysGlu 800

Qy        550 AAAGAAAAGTTGCAA-----------------GAACAACAAAGCGATTTAGAACAAGAG 591
                  ||||||:::::::::               |||::::::   :::   |||:::
Db        801 LysGluArgIleGluLeuGluLysGlnLysAlaGluGluGluGluGluAlaGluLysLeu 820

Qy        592 AGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTA---------------GAA 636
                  |||:::   ||||||:::   :::||||||:::   :::|||               |||
Db        821 ArgLysLeuLysGluGlnGluGluGluGlnLysLeuGluLeuLysArgIleArgAspGlu 840

Qy        637 CAAAGGAAGGCTGAT-----------------ACGAAAAAAAATTTAGAAAGAAAAAAG 678
                  |||:::|||   |||                 |||:::|||:::|||::::::|||
Db        841 GlnLysLysIleAspAspGluLysGluArgIleGluLysGluAsnIleGluLysGluLys 860

Qy        679 GAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCAGCTATAGAA 738
                  :::      :::         ::::::          |||||||||          :::
Db        861 GlnArgIleGluAsnGlnGluLysGlu---------ArgLeuGlu-----------Lys 873

Qy        739 CTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAATCTTCTTTACCTCAGGAC 798
                  |||      |||   :::|||   :::                           |||
Db        874 LeuGlnGlnGluGlnGlnArgLeuPhe----------------------------Asp 883

Qy        799 AACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA---ATAGAAAAAACAAATAGA 855
                  :::         :::   |||||||||:::   |||   :::   |||      |||
Db        884 GluGln------GluLysLysSerLysGluIleGluIleGlnLeuLeuLysGluGluArg 901

Qy        856 GAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACATAAAGGACATCTTGAAGAA 915
                  |||   :::                 :::       |||      |||
Db        902 GluGluLeu----------------------GluAlaGluLysAlaArgLeuAlaSer 913

Qy        916 AAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAATCTGCTGACATACAAAAT 975
                     |||      :::                 :::   :::   |||          :::
Db        914 LeuLysSerGluAlaAlaAlaAlaArgSerArgSerAsnSerSerThrSerAsnSerSer 933

Qy        976 CATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAG 1035
                  :::         :::   |||::::::      |||:::         :::
Db        934 AsnAsnSerSerSerSerAsnLeuAsnHis---AsnAsnThrAsnAsnAsnAsnAsnAsn 952

Qy       1036 GATGAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGATGATGAAGAC 1095
                  :::      :::      :::::::::      :::                ::::::
Db        953 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnThrSerThrAspAsn 972

Qy       1096 TTAGACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGA 1155
                  |||:::|||          |||
Db        973 LeuAsnGluSerAsnGlySerValThrProProProThrSerValSerArgArgGlnSer 992

Qy       1156 ATTTATAAAGAACTAGAAGATTTGATAGAGAAA-------------------------- 1188
                  |||         :::      |||::::::|||
Db        993 IleSerSerGlyIleGlySerLeuLeuLysLysAlaValProSerSerProSerSerSer 1012

Qy       1189 AATGAAAATTTAGATGATTTAGATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAA 1248
                  |||         :::   |||:::   :::   |||||||||      |||::::::
Db       1013 LysGluGluLysLysGluLysAspAspThrLeuAsnLysSerSerAlaSerLeuThrLys 1032

Qy       1249 GAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAGGATAATAATTTTAAACCAAAT 1308
                  |||:::   |||:::   :::|||   ||||||                  |||
Db       1033 GluGluLysLysGluLysGluLysIleGluLys-----------GluLeuLysGluLys 1048
```

```
Qy       1309 GATAAAAGTTTGTATGATGAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAG 1368
                  :::|||         :::|||      ||||||   |||   :::|||         |||
Db       1049 GluLysArgGluLysGluGluArgLysLysLysGluLysValGluLys---------Lys 1065

Qy       1369 GAAAAGGAAAAA 1380
                  ||||||||||||
Db       1066 GluLysGluLys 1069
```

```
RESULT 28
Q54C75_DICDI
ID   Q54C75_DICDI PRELIMINARY;      PRT;  2205 AA.
AC   Q54C75;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0191797;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000274; EAL60912.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   2205 AA;  254185 MW;  681BA55EC6E7C909 CRC64;

Alignment Scores:
Pred. No.:              2.17e-12        Length:         2205
Score:                  425.00          Matches:        146
Percent Similarity:     51.18%          Conservative:   136
Best Local Similarity:  26.50%          Mismatches:     129
Query Match:            17.09%          Indels:         140
DB:                     2               Gaps:           27

US-09-837-344-45 (1-1482) x Q54C75_DICDI (1-2205)

Qy          4 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAA 63
                :::::::::      ||||||:::|||:::      ||||||:::|||::::::::::
Db        610 LysGluLysLysLysLeuGluLysGluLysLeuGluLysGluArgLeuGluLysGluArg 629

Qy         64 AGCGAT---TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
                  :::   ||||||:::::::::|||   ||||||||||||:::::::::::
Db        630 LeuGluArgLeuGluLysGluLysLeuGluLysGluLysLeuGluLysGluLys------ 647

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATCTAGAACAA 180
                  ||||||:::|||:::|||   ||||||||||||::::::            ||||||:::
Db        648 LeuGluLysGluLysLeuGluLysGluLysLeuGluLys-----------LeuGluLys 663
```

```
Qy       181 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT---TTAGAACAAGAGAGA 237
Db       664 GluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluArgLeuGluLysGluArg 683

Qy       238 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCTAAA 297
Db       684 LeuGluLysGluLysLeuGluLysGluArg------LeuGluLysGluArgLeuGluLys 701

Qy       298 GAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTG 357
Db       702 GluArgLeuGluLysGluArg------LeuGluLysGluArgLeuGluLysGluArgLeu 719

Qy       358 CAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 417
Db       720 GluArg------------LeuGluLysGluArgLeuGluLysGluArgLeuGluLysGlu 735

Qy       418 CAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 477
Db       736 Arg------LeuGluLysGluArgValGluLysGluArgLeuGluLysGluArgGln--- 752

Qy       478 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---------GAT 528
Db       753 ---GluLysGluArgLeuGluLysGluArgLeuGluLysGluLysSerLeuArgGluGln 771

Qy       529 TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAA 588
Db       772 LeuGluLysGluArgLeuGluLysGluSerLeuGluLysGluArg------ValGluArg 789

Qy       589 GAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGGAAGGCT 648
Db       790 GluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluLysGluArgLeuGluLys 809

Qy       649 GATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAGGATTTA 708
Db       810 LeuLeuLysGluLysLeuGluLysGluLys----------------------SerLeu 821

Qy       709 TATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGATATTAT 768
Db       822 LysGluLysLeuGluLysGluLysIleGluLysAspLysGluLysGlu----------- 837

Qy       769 ATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAA 828
Db       838 --------------------------AspAspAspGluAsnGluAsnGluAspLys--- 847

Qy       829 ATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGG 888
Db       848 MetAspThrAspGlyLysValAsnArgSerSerLysThr---------GlnGluGlnArg 864

Qy       889 GATATA--------------CATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATA 933
Db       865 AspLeuGluAsnLeuLysSerArgGluGlnArgLysArgGluLysGluGluAspLeuMet 884

Qy       934 AAACCAGAACAAAAAGAAGATAAATCT----------------------GCTGACATA 969
Db       885 TyrTyrGluGlnAsnLysAsnLysSerAspGlnAsnArgGlyGlyGlyGlyGlyLeu 904

Qy       970 CAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTTCAAATAAGTAAG 1029
Db       905 ValSerGlnPheLeuGlnGlnSerAsnProLeuSerGlnLeuGlnLeuGlnSerAsnLys 924

Qy      1030 TATGAGGAT---GAAATAAGTGCTGAATATGACGATTCATTAATAGATGAAGAAGAAGAT 1086
Db       925 IleValAspLeuGluValIleGluAspTyr---------IleAlaGluArgLysArgGlu 941

Qy      1087 ------GATGAAGACTTAGACGAATTTAAGCCTATTGTG---------------------- 1119
Db       942 TrpPheGluLysAsnIleAspSerLeuLysLysMetAlaLeuProLeuTrpAsnLeuSer 961

Qy      1119 ------------------------------------------------------------ 1119

Db       962 ArgLysAsnMetGluValIleLysIleArgGluLysArgTyrPheGluTyrAsnAlaArg 981

Qy      1120 ---------CAATATGACAATTTCCAAGATGAAGAAAACATAGGAATTTATAAAGAACTA 1170

Db       982 LysCysLysLeuPheAsnAspPheGlnTyrGluPheSerProLysValTyrLysAlaIle 1001

Qy      1171 GAAGATTTGATAGAG---------AAAAATGAAAATTTAGATGATTTA------------ 1209
Db      1002 SerGlnThrPheGluAspPheTrpLysGluGluTrpIleValAspMetLeuHisSerGln 1021

Qy      1210 ---------GATGAAGGAATAGAAAAATCATCA-----------------------GAA 1236
Db      1022 LysProProProGluTyrValGluSerProSerGlnIleLysGlnGlnGlnGlnGlnGln 1041

Qy      1237 GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATATGAAAAAACAAAG--------- 1287
Db      1042 GlnGlnGlnGlnGlnGlnGlnGlnSerSerSerAlaGlnLysSerLysLeuIleVal 1061

Qy      1288 --------------------GATAATAATTTTAAACCAAATGAT--------------- 1311
Db      1062 ValGlnGlnGluThrAspAspAspAspAspAspSerProAspAspTrpAsnGluAspAsn 1081

Qy      1312 ---------------AAAAGTTTGTATGAT---------GAGCATATTAAAAAATATAAA 1347
Db      1082 LeuLeuSerGlnGlnSerSerLeuHisGluAsnLysGlyGluAsnGluAsnAspSerGlu 1101

Qy      1348 AATGATAAGCAGGTTAATAAGGAAAAGGAAAAA 1380
Db      1102 AsnGluSerGluSerGluSerGluLysGlnLys 1112
```

RESULT 29
Q54MJ1_DICDI
ID   Q54MJ1_DICDI PRELIMINARY;   PRT; 1061 AA.
AC   Q54MJ1;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0186744;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
CC   -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
DR   EMBL; AAFI01000127; EAL64490.1; -; Genomic_DNA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
KW   Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ   SEQUENCE   1061 AA; 128819 MW; 4B2361B2D1531D0D CRC64;

Alignment Scores:
Pred. No.:   3.9e-12   Length: 1061

```
Score:                  421.00      Matches:      146
Percent Similarity:     49.58%      Conservative: 147
Best Local Similarity:  24.70%      Mismatches:   178
Query Match:            16.93%      Indels:       120
DB:                     2           Gaps:         25

US-09-837-344-45 (1-1482) x Q54MJ1_DICDI (1-1061)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA--- 57
Db        266 LysGluGluGlnLeuLysArgAspGlnGluLysLeuGluArgGluLysLysArgGluTyr 285

Qy         58 ------------------------------------CAACAAAGCGATTTAGAACAA 78
Db        286 HisGlnGlnGlnProGlnGlnThrGlnIleGlnIleHisGlnGlnLysMetLeuGluLys 305

Qy         79 GATAGACTTGCTAAAGAAAAG------------TTACAAGAGCAGCAAAGCGATTTA--- 123
Db        306 GluArgLeuAlaGlnGluGluLysGluArgLeuLeuValGluGluArgGluArgLeuAsn 325

Qy        124 --------------------------GAACAAGAGAGACTTGCTAAAGAAAAGTTGCAA 156
Db        326 AlaThrArgValGluArgGluLysGlnGluLysGluArgLeuAspArgGluArgLysGlu 345

Qy        157 GAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAA---GAACAA 213
Db        346 ArgGluGlnAlaArgLeuGluLysGluArgLeuGluSerGluArgLeuGluLysGluArg 365

Qy        214 CAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 273
Db        366 GlnAlaArgValGluArgGluArgLysGluPheGluArgIleGluLysGluLysArgAsp 385

Qy        274 TTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAA 333
Db        386 ArgGluGlnThrArgLeuGluAsnGluArgLeuGluAsnGluArg------LeuLysLys 403

Qy        334 GAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGA 390
Db        404 GluLysLeuGluLysGluArgLeuGluAsnGluArgPheGluLysGluArgGlnGluArg 423

Qy        391 CGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAA---CAAGAGAGACTTGCT 447
Db        424 IleGluAsnGluArgLeuGluAsnGluArgPheGluArgGluArgLysGluArgLeuGlu 443

Qy        448 AAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 507
Db        444 LysGluLysIleGluLysGluGluArgGlu---AsnGlnLeuArgLeuAlaLysGluLys 462

Qy        508 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 567
Db        463 ---GluGluLysGluGluLysGluArgGlnGluArgGlnLeuLysGluArgLeu---Glu 480

Qy        568 CAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGA 627
Db        481 ArgLysGluLysLeuGluLysGluArgLysGluArgGluGluLysGluGluLysAspArg 500

Qy        628 GATTTAGAACAAAGGAAGGCTGATACGAAAAAAAATTTAGAAAGA--------------- 672
Db        501 ---------GlnGluArgLeuGlnLeuLysGluArgLeuAspArgGluArgLysGluArg 517

Qy        673 ---AAAAAGGAACATGGAGATATATTAGCAGAGGATTTATATGGTCGTTTAGAAATACCA 729
Db        518 LeuGluLysGluArgLysGluArgGluGluLysAspGluLysAspArgGlnGluArgLeu 537

Qy        730 GCTATAGAACTTCCATCAGAAAATGAACGTGGATATTATATACCACATCAA--------- 780
Db        538 GlnIleLysGluArgLeuGluArgGluArgLysGluArgLeuGluLysGluArgLeuGlu 557

Qy        781 ---TCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCCAAGGAAATGTCTATA 837
Db        558 TyrGluArgLeuGluThrLeuArgLysGluLysGluLysGluArgLysGluLysPheGlu 577

Qy        838 ATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGACGAAGGGATATACAT 897
Db        578 ThrAlaArgLeuGluArgGluLysLeuGluLeuGluLeuLysGluLysAspGluLysLys 597

Qy        898 AAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAACAAAAAGAAGATAAA 957
Db        598 Arg---------GluGluLysArgLeuAlaLysIleIleGluGluAsnArgLysLysGlu 614

Qy        958 TCTGCTGACATACAAAATCATACATTAGAGACAGTAAATATTTCTGATGTTAATGATTTT 1017
Db        615 LysAlaLeuValAspLysLysLeuGluGluGlnGluArgLeuSer------HisGluLeu 632

Qy       1018 CAAATAAGTAAGTATGAG--------------GATGAAATAAGTGCTGAATATGACGAT 1062
Db        633 LysIleGlnLysGluLysGluIleLeuLysMetGluLysLeuAspLeuGlnCysGluLys 652

Qy       1063 TCATTAATAGATGAAGAAGAAGATGATGAAGACTTAGAC---------------------- 1101
Db        653 LysLysLeuAspLysGluLysLysGluLeuGlnIleGluLysGluLeuSerGlnLysGln 672

Qy       1102 ---GAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
Db        673 LeuGlnPheIleGluLeuLysGlnGlnGluIleAsnGlnGlnGlnGlnGluIleAsnHis 692

Qy       1159 TATAAAGAACTAGAAGAT--------------------TTGATAGAGAAAAATGAAAAT 1197
Db        693 GlnGlnGlnGlnGlnGlnGlnThrGlnGlnGlnIleProLeuLeuThrLysGlnProThr 712

Qy       1198 TTAGATGATTTA---------------------------------------------- 1209
Db        713 ValGluGluIleGluTyrGluArgValGlnGluGluIleLysGlnLysLysLeuGluArg 732

Qy       1210 GATGAAGGAATAGAAAAATCATCAGAAGAATTATCTGAAGAAAAAATAAAAAAAGGAAAG 1269
Db        733 GluGlnArgLeuLysGlnGlnLeuGluGluGluGluLysGluArgLeuLysArgGlnGlu 752

Qy       1270 AAATATGAAAAAACAAAGGATAATAATTTTAAACCAAATGATAAAAGTTTGTATGATGAG 1329
Db        753 GluArgGlnLysArgLysGluGln------------ArgAspLysGluLeuLysGluGlu 768

Qy       1330 CATATT---------------------------AAAAAATATAAAAATGATAAG------ 1356
Db        769 GluLeuArgIleGluAlaGluArgGluThrArgLysLysLeuLysGluGluArgGluLeu 788

Qy       1357 CAGGTTAATAAGGAA---------------------------------AAGGAAAAA 1380
Db        789 GlnPheLysLysGluGlnValGlnLeuLeuLeuGluArgAspLysArgValLysGluLys 808

Qy       1381 TTCATAAAATCATTGTTTCATATATTTGAC---GGAGACAATGAAATTTTACAGATCGTG 1437
Db        809 GlnLeuLysAspGlnLysTyrThrTyrLysIleAsnAspThrGluLeuGluHisIleVal 828

Qy       1438 GATGAGTTATCTGAAGATATAACTAAATATTTT 1470
Db        829 GluAsp------LysAspLeuThrLysGluPhe 837

RESULT 30
Q50RA3_ENTHI
ID   Q50RA3_ENTHI PRELIMINARY;      PRT;   762 AA.
AC   Q50RA3;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=295.t00001;
OS   Entamoeba histolytica HM-1:IMSS.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=294381;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=HM-1:IMSS;
RX   PubMed=15729342; DOI=10.1038/nature03291;
RA   Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA   Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
```

```
RA   Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA   Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA   Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA   Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA   Quail M.A., Rabbinowitsch E., Norbertczak H., Price C., Wang Z.,
RA   Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA   Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA   El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA   Fraser C.M., Hall N.;
RT   "The genome of the protist parasite Entamoeba histolytica.";
RL   Nature 433:865-868(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFB01000803; EAL44122.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   762 AA;  94161 MW;  A406F044AB7E1F84 CRC64;

Alignment Scores:
Pred. No.:              5.18e-12      Length:         762
Score:                  419.00        Matches:        130
Percent Similarity:     48.95%        Conservative:   103
Best Local Similarity:  27.31%        Mismatches:     149
Query Match:            16.85%        Indels:         94
DB:                     2             Gaps:           21

US-09-837-344-45 (1-1482) x Q50RA3_ENTHI (1-762)

Qy          1 CAAGAACAACAAAGCGATCTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAA 60
              :::   |||:::   :::::::::::::::         |||   ||||||:::|||
Db        109 LysSerGlnGluGluGluValGlnLysGlnGluGluGluLysArgLysLeuGluGluIle 128

Qy         61 CAAAGCGATTTAGAACAAGATAGACTTGCTAAAGAAAAGTTACAAGAGCAGCAAAGCGAT 120
              |||       |||:::           |||   ||||||:::|||::::::
Db        129 GlnArgGlnGluGluGlu-----------LysArgLysLeuGluGluGluGluGluLys 144

Qy        121 TTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC---GATCTAGAA 177
              ::::::   |||         |||   :::|||:::|||::::::         ||||||
Db        145 MetLysLeuGluGluGluGluLysAsnArgLeuGluGluGluGluLysArgLysLeuGlu 164

Qy        178 CAAGAG---AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGAG 234
              :::|||   ::::::   ||||||:::::::|||   |||:::   :::::::::::::::
Db        165 GluGluIleGlnLysGlnLysGluGluIleGlnArgGlnGluGluGluIleGlnLysGln 184

Qy        235 AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTAGAACAAGATAGACTTGCT 294
                  |||   ||||||:::|||:::::::::   ||||||
Db        185 GluGluGluLysMetLysLeuGluGluGluLysAsnArgLeuGlu--------------- 199

Qy        295 AAAGAAAAGTTACAAGAGCAGCAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAG 354
                ||:::::|||:::|||:::   :::   :::::::::         |||   |||
Db        200 ---GluGluIleGlnLysGlnGluGluGluArgGlnLysGlnGluGluGluLysAsnLys 218

Qy        355 TTGCAAGAACAACAAAGCGATTTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAA 414
              |||:::|||:::             |||:::|||:::|||         ||||||:::|||
Db        219 LeuGluGluGluIleGlnLysGlnGluGluGluLysArg---------LysLeuGluGlu 235

Qy        415 CAACAAAGCGATTTAGAACAAGAGAGACTTGCTAAAGAAAAGTTGCAAGAACAACAAAGC 474
              ::::::      ||||||:::   :::      ||||||::::::|||:::|||:::
Db        236 GluLysLysArgLeuGluGluThrLysArgGlnLysGluGluIleGlnLysGlnGluGlu 255

Qy        475 GATTTAGAACAAGAG------AGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGAT 528
                  :::||||||      :::|||   ||||||::::::|||   |||:::   :::
Db        256 ArgGlnLysGlnGluGluIleGlnArgGlnLysGluGluIleGlnArgGlnGluGluGlu 275

Qy        529 ------TTAGAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAACAACAAAGCGATTTA 582
                    ||||||         |||   |||   :::|||   |||
Db        276 GluArgLysArgGlnGluGluGluArgLysArgLysGluGluGluIleGlnArgGlnLys 295

Qy        583 GAACAAGAGAGACGTGCTAAAGAAAAGTTGCAAGAGCAGCAAAGAGATTTAGAACAAAGG 642
              |||:::||||||:::   :::|||::::::|||   |||:::         |||:::|||
Db        296 GluGluGluArgLysLysGlnGluGluIleGlnArgGlnLys---------GluGluArg 312

Qy        643 AAGGCTGATACGAAAAAAAATTTAGAAAGAAAAAAGGAACATGGAGATATATTAGCAGAG 702
                |||         |||:::      |||||||||::::::
Db        313 Lys------ArgLysGluGluGluGluArgLysArgGln---------------------- 323

Qy        703 GATTTATATGGTCGTTTAGAAATACCAGCTATAGAACTTCCATCAGAAAATGAACGTGGA 762
                                                          |||   ||||||
Db        324 ------------------------------------------GluGluGluArg--- 327

Qy        763 TATTATATACCACATCAATCTTCTTTACCTCAGGACAACAGAGGGAATAGTAGAGATTCC 822
                                    |||                   |||
Db        328 ----------------------------GlnLysGlnGluGluIleGlnArgGlnLys 337

Qy        823 AAGGAAATGTCTATAATAGAAAAAACAAATAGAGAATCTATTACAACAAATGTTGAAGGA 882
              :::|||         |||::::::         |||   |||            |||
Db        338 GluGlu---------IleGlnArgGlnLysGluGluGluIleGlnArgGlnLysGluGlu 354

Qy        883 CGAAGGGATATACATAAAGGACATCTTGAAGAAAAGAAAGATGGTTCAATAAAACCAGAA 942
              |||:::         |||            |||||||||:::      :::|||   :::
Db        355 ArgLys------ArgLysGluGluGluArgGluLysLysGluArgGluLeuLysArgLys 372

Qy        943 CAAAAAGAAGATAAATCTGCT---------GACATACAAAATCATACATTAGAGACAGTA 993
                 :::||||||:::               |||   :::|||         |||   :::
Db        373 LeuGluGluAspArgLysValLeuGluGluAspLysGluAsnGlnLysAspGluLysIle 392

Qy        994 AATATTTCTGATGTTAATGATTTTCAAATAAGTAAGTATGAGGATGAAATAAGTGCTGAA 1053
              |||   |||            :::|||:::      |||||||||:::
Db        393 IleIleLysAsp-----------GluIleAsnIleLeuGluAspGluVal--------- 405

Qy       1054 TATGACGATTCATTAATAGATGAA--------------GAAGAAGATGATGAAGACTTA 1098
                           |||::::::|||                |||||:::   :::   :::
Db        406 ---------HisLeuLeuGluGluLysLysGlyGluLeuGluGluGluLysGlnArgIle 422

Qy       1099 GACGAATTTAAGCCTATTGTGCAATATGACAATTTCCAAGATGAAGAAAACATAGGAATT 1158
              ||||||            :::      |||::::::|||:::
Db        423 AspGlu--------------GluMetHisLysPheGluGluGluLys------------ 433

Qy       1159 TATAAAGAACTAGAAGATTTGATAGAGAAAAATGAAAATTTAGATGATTTAGATGAAGGA 1218
                 |||   :::|||:::   |||
Db        434 ---LysArgIleGluGluGluIleAsnAsnLysArgGlyAlaSerAlaSerArgMetTyr 452

Qy       1219 ATAGAAAAATCATCAGAA---GAATTATCTGAAGAAAAAATAAAAAAAGGAAAGAAATAT 1275
                    |||       |||   ||||||   :::   ||||||:::      :::|||
Db        453 PheAlaLysLysGlnGluIleGluLeuGlnGlnAsnLysIleGluAsnLysGlnLysGlu 472

Qy       1276 GAAAAAACAAAGGATAATAATTTTAAACCAAATGAT---------AAAAGTTTGTATGAT 1326
                    |||               |||   |||              |||      :::
Db        473 AlaThrThrAsnLysMetLysIleLysLysAsnGlnSerGluIleLysArgLysGlnGlu 492

Qy       1327 GAGCATATTAAAAAATATAAAAATGATAAGCAGGTTAATAAGGAAAAG 1374
              |||   |||||||||   |||      |||:::::::::|||   |||
Db        493 GluIleIleLysLysGluLysGluIleLysGluIleAspLysArgLys 508

Search completed: December 21, 2005, 11:27:08
Job time : 421.484 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 2.0261 Seconds
(without alignments)
5204.614 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4883371

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22895/app_query.fasta_1.3740
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_821_@runat_21122005_093215_22895 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 10:46:06
Job time : 2.0261 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:22 ; Search time 0.415663 Seconds
(without alignments)
4773.622 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1143330

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093212_22815/app_query.fasta_1.3740
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_161_@runat_21122005_093212_22815 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 09:38:26
Job time : 0.415663 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:32:23 ; Search time 1.89759 Seconds
(without alignments)
5284.546 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3732626

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22847/app_query.fasta_1.3740
-DB=Published_Applications_AA_Main -QFMT=fastan -SUFFIX=rapbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_781_@runat_21122005_093214_22847
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 11:01:57
Job time : 1.89759 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 09:38:31 ; Search time 0.0682731 Seconds
(without alignments)
2507.145 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 53982 seqs, 7132107 residues

Total number of hits satisfying chosen parameters: 107923

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093214_22867/app_query.fasta_1.3740
-DB=Published_Applications_AA_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09837344_@CGN_1_1_2_@runat_21122005_093214_22867
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	20	95.2	9	7	US-11-097-864-11	Sequence 11, Appl
c 2	20	95.2	9	7	US-11-097-864-150	Sequence 150, App
c 3	20	95.2	9	7	US-11-097-864-353	Sequence 353, App
c 4	20	95.2	9	7	US-11-097-864-463	Sequence 463, App
c 5	20	95.2	9	7	US-11-097-864-587	Sequence 587, App
c 6	20	95.2	9	7	US-11-097-912-11	Sequence 11, Appl
c 7	20	95.2	9	7	US-11-097-912-150	Sequence 150, App
c 8	20	95.2	9	7	US-11-097-912-353	Sequence 353, App
c 9	20	95.2	9	7	US-11-097-912-463	Sequence 463, App
c 10	20	95.2	9	7	US-11-097-912-587	Sequence 587, App
c 11	20	95.2	10	7	US-11-097-864-337	Sequence 337, App
c 12	20	95.2	10	7	US-11-097-864-509	Sequence 509, App
c 13	20	95.2	10	7	US-11-097-912-337	Sequence 337, App
c 14	20	95.2	10	7	US-11-097-912-509	Sequence 509, App
c 15	20	95.2	46	6	US-10-467-657-8088	Sequence 8088, Ap
16	20	100.0	71	7	US-11-132-142-17	Sequence 17, Appl
17	20	100.0	115	7	US-11-156-909-11	Sequence 11, Appl
18	20	100.0	133	7	US-11-156-909-10	Sequence 10, Appl
c 19	20	95.2	135	6	US-10-467-657-2028	Sequence 2028, Ap
c 20	20	95.2	136	6	US-10-793-626-912	Sequence 912, App
21	20	100.0	149	7	US-11-156-909-31	Sequence 31, Appl
c 22	20	95.2	166	6	US-10-467-657-3512	Sequence 3512, Ap
23	20	100.0	169	6	US-10-793-626-1534	Sequence 1534, Ap
24	20	100.0	173	7	US-11-156-909-29	Sequence 29, Appl
c 25	20	95.2	176	6	US-10-793-626-362	Sequence 362, App
26	20	100.0	184	6	US-10-467-657-5748	Sequence 5748, Ap
c 27	20	95.2	208	7	US-11-055-822-462	Sequence 462, App
c 28	20	95.2	208	7	US-11-055-822-704	Sequence 704, App
29	20	100.0	212	6	US-10-821-234-1430	Sequence 1430, Ap
c 30	20	95.2	223	6	US-10-878-556A-167	Sequence 167, App
31	20	100.0	230	6	US-10-793-626-3242	Sequence 3242, Ap
32	20	100.0	242	7	US-11-100-183-12	Sequence 12, Appl
33	20	100.0	249	6	US-10-467-657-2256	Sequence 2256, Ap
34	20	100.0	259	7	US-11-000-463-276	Sequence 276, App
c 35	20	95.2	266	6	US-10-467-657-5258	Sequence 5258, Ap
36	20	100.0	285	7	US-11-179-843-3	Sequence 3, Appli
c 37	20	95.2	287	6	US-10-467-657-2960	Sequence 2960, Ap
38	20	100.0	299	6	US-10-131-826A-366	Sequence 366, App
39	20	100.0	299	7	US-11-000-463-275	Sequence 275, App
40	20	100.0	299	7	US-11-000-463-747	Sequence 747, App
41	20	100.0	299	7	US-11-000-463-748	Sequence 748, App
42	20	100.0	299	7	US-11-173-037-7	Sequence 7, Appli
c 43	20	95.2	299	7	US-11-112-882-30	Sequence 30, Appl
44	20	100.0	300	7	US-11-025-834A-21	Sequence 21, Appl
45	20	100.0	316	6	US-10-793-626-1350	Sequence 1350, Ap

ALIGNMENTS

RESULT 1
US-11-097-864-11
; Sequence 11, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006205
; CURRENT APPLICATION NUMBER: US/11/097,864
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-864-11

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0

```
DB:                    7             Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-864-11 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          4 ValTyrAsnLeu 7

RESULT 2
US-11-097-864-150
; Sequence 150, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-150

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-864-150 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          6 ValTyrAsnLeu 9

RESULT 3
US-11-097-864-353
; Sequence 353, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 353
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-353

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-864-353 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 4
US-11-097-864-463
; Sequence 463, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 463
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-463

Alignment Scores:
Pred. No.:              1.56e+06       Length:        9
Score:                  20.00          Matches:       3
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  75.00%         Mismatches:    0
Query Match:            95.24%         Indels:        0
DB:                     7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-864-463 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5

RESULT 5
US-11-097-864-587
; Sequence 587, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
```

; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006205
; CURRENT APPLICATION NUMBER: US/11/097,864
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 587
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-864-587

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-864-587 (1-9)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          5 ValTyrAsnLeu 8
```

RESULT 6
US-11-097-912-11
; Sequence 11, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006204
; CURRENT APPLICATION NUMBER: US/11/097,912
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-912-11

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-912-11 (1-9)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          4 ValTyrAsnLeu 7
```

RESULT 7
US-11-097-912-150
; Sequence 150, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006204
; CURRENT APPLICATION NUMBER: US/11/097,912
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-912-150

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-912-150 (1-9)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          6 ValTyrAsnLeu 9
```

RESULT 8
US-11-097-912-353
; Sequence 353, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006204
; CURRENT APPLICATION NUMBER: US/11/097,912
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 353
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-912-353

Alignment Scores:

Pred. No.:	1.56e+06	Length:	9
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1

```
Best Local Similarity:  75.00%        Mismatches:    0
Query Match:            95.24%        Indels:        0
DB:                     7             Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-912-353 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 9
US-11-097-912-463
; Sequence 463, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 463
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-463

Alignment Scores:
Pred. No.:              1.56e+06      Length:        9
Score:                  20.00         Matches:       3
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  75.00%        Mismatches:    0
Query Match:            95.24%        Indels:        0
DB:                     7             Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-912-463 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5

RESULT 10
US-11-097-912-587
; Sequence 587, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006204
;  CURRENT APPLICATION NUMBER: US/11/097,912
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 587
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-912-587

Alignment Scores:
Pred. No.:              1.56e+06      Length:        9
Score:                  20.00         Matches:       3
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  75.00%        Mismatches:    0
Query Match:            95.24%        Indels:        0
DB:                     7             Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-912-587 (1-9)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          5 ValTyrAsnLeu 8

RESULT 11
US-11-097-864-337
; Sequence 337, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
;  APPLICANT:  Jakobovits, Aya
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
;  TITLE OF INVENTION:  USEFUL IN THE TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 511582006205
;  CURRENT APPLICATION NUMBER: US/11/097,864
;  CURRENT FILING DATE:  2005-04-01
;  PRIOR APPLICATION NUMBER: US 10/062,109
;  PRIOR FILING DATE: 2002-01-31
;  PRIOR APPLICATION NUMBER: US 10/005,480
;  PRIOR FILING DATE: 2001-11-07
;  NUMBER OF SEQ ID NOS: 765
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 337
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-11-097-864-337

Alignment Scores:
Pred. No.:              200           Length:        10
Score:                  20.00         Matches:       3
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  75.00%        Mismatches:    0
Query Match:            95.24%        Indels:        0
DB:                     7             Gaps:          0

US-09-837-344-46 (1-12) x US-11-097-864-337 (1-10)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 12
US-11-097-864-509
; Sequence 509, Application US/11097864
; Publication No. US20050265924A1
; GENERAL INFORMATION:
;  APPLICANT: Challita-Eid, Pia M.
;  APPLICANT:  Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Morrison, Karen Jane Meyrick
```

; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN NAMED 161P2F10B
; TITLE OF INVENTION: USEFUL IN THE TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006205
; CURRENT APPLICATION NUMBER: US/11/097,864
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-864-509

Alignment Scores:

Pred. No.:	200	Length:	10
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-864-509 (1-10)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5
```

RESULT 13
US-11-097-912-337
; Sequence 337, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006204
; CURRENT APPLICATION NUMBER: US/11/097,912
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 337
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-912-337

Alignment Scores:

Pred. No.:	200	Length:	10
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-912-337 (1-10)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          1 ValTyrAsnLeu 4
```

RESULT 14
US-11-097-912-509
; Sequence 509, Application US/11097912
; Publication No. US20050265921A1
; GENERAL INFORMATION:
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 161P2F10B
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 511582006204
; CURRENT APPLICATION NUMBER: US/11/097,912
; CURRENT FILING DATE: 2005-04-01
; PRIOR APPLICATION NUMBER: US 10/062,109
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-097-912-509

Alignment Scores:

Pred. No.:	200	Length:	10
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	7	Gaps:	0

US-09-837-344-46 (1-12) x US-11-097-912-509 (1-10)

```
Qy         12 ATATATAACCTT 1
              :::|||||||||
Db          2 ValTyrAsnLeu 5
```

RESULT 15
US-10-467-657-8088
; Sequence 8088, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8088
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8088

Alignment Scores:

Pred. No.:	250	Length:	46
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0

```
DB:                    6              Gaps:          0

US-09-837-344-46 (1-12) x US-10-467-657-8088 (1-46)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db         11 ValTyrAsnLeu 14

RESULT 16
US-11-132-142-17
; Sequence 17, Application US/11132142
; Publication No. US20050276818A1
; GENERAL INFORMATION:
;  APPLICANT: The Burnham Institute
;  APPLICANT:  Sikora, Sergey
;  APPLICANT:  Godzik, Adam
;  TITLE OF INVENTION: UNCHARACTERIZED ORF3 IN SARS-CORONAVIRUS IS A
;  TITLE OF INVENTION:  CYCLIC-AMP-DEPENDENT KINASE AND A TARGET FOR SARS THERAPY
;  FILE REFERENCE: 8014-011-US
;  CURRENT APPLICATION NUMBER: US/11/132,142
;  CURRENT FILING DATE:  2005-05-17
;  PRIOR APPLICATION NUMBER: 60/571,698
;  PRIOR FILING DATE: 2004-05-17
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 17
;   LENGTH: 71
;   TYPE: PRT
;   ORGANISM: Arabidopsis sp.
US-11-132-142-17

Alignment Scores:
Pred. No.:             267            Length:        71
Score:                 20.00          Matches:       4
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-132-142-17 (1-71)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         36 LysValIleTyr 39

RESULT 17
US-11-156-909-11
; Sequence 11, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 11
;   LENGTH: 115
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Mature Protein for Clone 3.4
US-11-156-909-11

Alignment Scores:
Pred. No.:             286            Length:        115
Score:                 20.00          Matches:       4
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-156-909-11 (1-115)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        102 LysValIleTyr 105

RESULT 18
US-11-156-909-10
; Sequence 10, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 10
;   LENGTH: 133
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Precursor Protein for Clone 3.4
US-11-156-909-10

Alignment Scores:
Pred. No.:             292            Length:        133
Score:                 20.00          Matches:       4
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    7              Gaps:          0

US-09-837-344-46 (1-12) x US-11-156-909-10 (1-133)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        120 LysValIleTyr 123

RESULT 19
US-10-467-657-2028
; Sequence 2028, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2028
;   LENGTH: 135
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2028

Alignment Scores:
Pred. No.:             293            Length:        135
Score:                 20.00          Matches:       3
```

Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 95.24% Indels: 0
DB: 6 Gaps: 0

US-09-837-344-46 (1-12) x US-10-467-657-2028 (1-135)

```
Qy          12 ATATATAACCTT 1
               :::|||||||||
Db          85 ValTyrAsnLeu 88
```

RESULT 20
US-10-793-626-912
; Sequence 912, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 912
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-912

Alignment Scores:
Pred. No.: 293 Length: 136
Score: 20.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 95.24% Indels: 0
DB: 6 Gaps: 0

US-09-837-344-46 (1-12) x US-10-793-626-912 (1-136)

```
Qy          12 ATATATAACCTT 1
               :::|||||||||
Db         123 ValTyrAsnLeu 126
```

RESULT 21
US-11-156-909-31
; Sequence 31, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
; APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.
; FILE REFERENCE: RB-125-SEQ D3
; CURRENT APPLICATION NUMBER: US/11/156,909
; CURRENT FILING DATE: 2005-06-20
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2004-08-12
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 31
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Tenebrio molitor
; FEATURE:
; OTHER INFORMATION: Mature Protein with His-tag, Clone 3.4
US-11-156-909-31

Alignment Scores:
Pred. No.: 297 Length: 149
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-837-344-46 (1-12) x US-11-156-909-31 (1-149)

```
Qy           1 AAGGTTATATAT 12
               ||||||||||||
Db         136 LysValIleTyr 139
```

RESULT 22
US-10-467-657-3512
; Sequence 3512, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3512
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3512

Alignment Scores:
Pred. No.: 302 Length: 166
Score: 20.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 95.24% Indels: 0
DB: 6 Gaps: 0

US-09-837-344-46 (1-12) x US-10-467-657-3512 (1-166)

```
Qy          12 ATATATAACCTT 1
               :::|||||||||
Db           1 ValTyrAsnLeu 4
```

RESULT 23
US-10-793-626-1534
; Sequence 1534, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1534
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1534

```
Alignment Scores:
Pred. No.:                303             Length:         169
Score:                    20.00           Matches:        4
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-837-344-46 (1-12) x US-10-793-626-1534 (1-169)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         57 LysValIleTyr 60

RESULT 24
US-11-156-909-29
; Sequence 29, Application US/11156909
; Publication No. US20050274924A1
; GENERAL INFORMATION:
;  APPLICANT: Horwath, K. L. ,Easton, C. M. , and Myers, K. L.
;  TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
;  TITLE OF INVENTION:  Antifreeze Proteins and Method for Assaying Activity.
;  FILE REFERENCE: RB-125-SEQ D3
;  CURRENT APPLICATION NUMBER: US/11/156,909
;  CURRENT FILING DATE:  2005-06-20
;  PRIOR APPLICATION NUMBER: 60/210,446
;  PRIOR FILING DATE: 2004-08-12
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: Microsoft Word
; SEQ ID NO 29
;   LENGTH: 173
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   OTHER INFORMATION: Precursor protein with His-tag, Clone 3.4
US-11-156-909-29

Alignment Scores:
Pred. No.:                304             Length:         173
Score:                    20.00           Matches:        4
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-837-344-46 (1-12) x US-11-156-909-29 (1-173)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        160 LysValIleTyr 163

RESULT 25
US-10-793-626-362
; Sequence 362, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 362
;   LENGTH: 176
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-10-793-626-362

Alignment Scores:
Pred. No.:                305             Length:         176
Score:                    20.00           Matches:        3
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    75.00%          Mismatches:     0
Query Match:              95.24%          Indels:         0
DB:                       6               Gaps:           0

US-09-837-344-46 (1-12) x US-10-793-626-362 (1-176)

Qy         12 ATATATAACCTT 1
                :::|||||||||
Db          1 ValTyrAsnLeu 4

RESULT 26
US-10-467-657-5748
; Sequence 5748, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: FONTANA Maria Rita
;  APPLICANT: PIZZA Mariagrazia
;  APPLICANT: MASIGNANI Vega
;  APPLICANT: MONACI Elisabetta
;  TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/467,657
;  CURRENT FILING DATE:  2003-08-11
;  PRIOR APPLICATION NUMBER: GB-0103424.8
;  PRIOR FILING DATE: 2001-02-12
;  NUMBER OF SEQ ID NOS: 9218
;  SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5748
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5748

Alignment Scores:
Pred. No.:                307             Length:         184
Score:                    20.00           Matches:        4
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-837-344-46 (1-12) x US-10-467-657-5748 (1-184)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        177 LysValIleTyr 180

RESULT 27
US-11-055-822-462
; Sequence 462, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
```

```
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 462
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-462

Alignment Scores:
Pred. No.:              312          Length:         208
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
DB:                     7            Gaps:           0

US-09-837-344-46 (1-12) x US-11-055-822-462 (1-208)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db         14 ValTyrAsnLeu 17

RESULT 28
US-11-055-822-704
; Sequence 704, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
;  TITLE OF INVENTION:  METABOLIC PATHWAY PROTEINS
;  FILE REFERENCE: BGI-121CPCN
;  CURRENT APPLICATION NUMBER: US/11/055,822
;  CURRENT FILING DATE:  2005-02-11
;  PRIOR APPLICATION NUMBER: 09/606,740
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: 60/141,031
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/142,101
;  PRIOR FILING DATE: 1999-07-02
;  PRIOR APPLICATION NUMBER: 60/148,613
;  PRIOR FILING DATE: 1999-08-12
;  PRIOR APPLICATION NUMBER: 60/187,970
;  PRIOR FILING DATE: 2000-03-09
;  PRIOR APPLICATION NUMBER: DE 19930476.9
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: DE 19931415.2
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931418.7
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931419.5
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: DE 19931420.9
;  PRIOR FILING DATE: 1999-07-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 704
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-704

Alignment Scores:
Pred. No.:              312          Length:         208
Score:                  20.00        Matches:        3
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  75.00%       Mismatches:     0
Query Match:            95.24%       Indels:         0
DB:                     7            Gaps:           0

US-09-837-344-46 (1-12) x US-11-055-822-704 (1-208)

Qy         12 ATATATAACCTT 1
              :::|||||||||
Db         14 ValTyrAsnLeu 17

RESULT 29
US-10-821-234-1430
; Sequence 1430, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1430
;   LENGTH: 212
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1430

Alignment Scores:
Pred. No.:              313          Length:         212
Score:                  20.00        Matches:        4
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            100.00%      Indels:         0
DB:                     6            Gaps:           0

US-09-837-344-46 (1-12) x US-10-821-234-1430 (1-212)

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         64 LysValIleTyr 67

RESULT 30
US-10-878-556A-167
; Sequence 167, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 167
;   LENGTH: 223
;   TYPE: PRT
```

; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: sw_hum/kad4_human
; DATABASE ENTRY DATE: 1992-08-01
US-10-878-556A-167

Alignment Scores:

Pred. No.:	315	Length:	223
Score:	20.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	0
Query Match:	95.24%	Indels:	0
DB:	6	Gaps:	0

US-09-837-344-46 (1-12) x US-10-878-556A-167 (1-223)

```
Qy          12 ATATATAACCTT 1
               :::|||||||||
Db         135 ValTyrAsnLeu 138
```

Search completed: December 21, 2005, 11:03:39
Job time : 1.06827 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:27:59 ; Search time 69.5382 Seconds
(without alignments)
9809.312 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	12	100.0	12	6	AR162971	AR162971 Sequence
2	12	100.0	12	6	AR162975	AR162975 Sequence
3	12	100.0	12	6	AR178260	AR178260 Sequence
4	12	100.0	12	6	AR178264	AR178264 Sequence
5	12	100.0	23	6	A36617	A36617 Sequence 26
6	12	100.0	104	10	G32838	G32838 A009U15 Hum
c 7	12	100.0	113	10	CR378386	CR378386 Arabidops
8	12	100.0	139	10	BX294638	BX294638 Arabidops
c 9	12	100.0	146	6	BD043274	BD043274 Sequence
c 10	12	100.0	146	6	AX907741	AX907741 Sequence
c 11	12	100.0	169	2	AF487987	AF487987 Triatoma
c 12	12	100.0	184	6	BD268181	BD268181 Gene to b
c 13	12	100.0	184	6	AR210010	AR210010 Sequence
14	12	100.0	188	6	AX196591	AX196591 Sequence
15	12	100.0	190	10	G23922	G23922 human STS W
c 16	12	100.0	200	10	AB153981	AB153981 Homo sapi
17	12	100.0	201	15	ATH531582	AJ531582 Arabidops
18	12	100.0	201	15	ATH531598	AJ531598 Arabidops
19	12	100.0	225	10	G31833	G31833 SpAGH1 Norw
c 20	12	100.0	226	15	AY020751	AY020751 Oryza sat
21	12	100.0	227	9	MMU409475	AJ409475 Mus muscu
c 22	12	100.0	232	10	AB143702	AB143702 Homo sapi
23	12	100.0	232	15	AY020956	AY020956 Oryza sat
24	12	100.0	234	13	AY114172	AY114172 Human her
c 25	12	100.0	236	10	G37836	G37836 wh_hsp1 Pla
c 26	12	100.0	236	10	G44501	G44501 C9M106 Plas
c 27	12	100.0	238	10	AU026951	AU026951 Rattus no
c 28	12	100.0	242	6	BD028743	BD028743 Sequence
c 29	12	100.0	242	6	AX893210	AX893210 Sequence
30	12	100.0	255	6	CQ457358	CQ457358 Sequence
31	12	100.0	255	10	G11552	G11552 human STS S
32	12	100.0	255	10	G26557	G26557 human STS S
c 33	12	100.0	264	10	BV344514	BV344514 S230P6149
34	12	100.0	267	6	AX311604	AX311604 Sequence
c 35	12	100.0	277	6	CQ475242	CQ475242 Sequence
c 36	12	100.0	281	10	AB152907	AB152907 Homo sapi
37	12	100.0	288	6	CQ748906	CQ748906 Sequence
c 38	12	100.0	289	6	CQ457286	CQ457286 Sequence
c 39	12	100.0	293	4	AF080181	AF080181 Procyon l
c 40	12	100.0	294	6	BD042247	BD042247 Sequence
c 41	12	100.0	294	6	AX906714	AX906714 Sequence
c 42	12	100.0	295	4	AF080174	AF080174 Procyon l
c 43	12	100.0	295	4	AF080175	AF080175 Procyon l
c 44	12	100.0	295	4	AF080178	AF080178 Procyon l
c 45	12	100.0	295	4	AF080180	AF080180 Procyon l

ALIGNMENTS

RESULT 1
AR162971
LOCUS AR162971 12 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 42 from patent US 6270771.
ACCESSION AR162971
VERSION AR162971.1 GI:16233436
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 12)
AUTHORS Guerin-Marchand,C. and Druilhe,P.
TITLE Peptide sequences specific for the hepatic stages of P. falciparum bearing epitopes capable of stimulating the T lymphocytes
JOURNAL Patent: US 6270771-A 42 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .12
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 8.3e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 1 AAGGTTATATAT 12

RESULT 2
AR162975
LOCUS AR162975 12 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 46 from patent US 6270771.
ACCESSION AR162975
VERSION AR162975.1 GI:16233442
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 12)

```
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6270771-A 46 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
AR178260
LOCUS       AR178260                  12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 42 from patent US 6319502.
ACCESSION   AR178260
VERSION     AR178260.1  GI:20219398
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 42 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
AR178264
LOCUS       AR178264                  12 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 46 from patent US 6319502.
ACCESSION   AR178264
VERSION     AR178264.1  GI:20219402
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 12)
  AUTHORS   Guerin-Marchand,C. and Druilhe,P.
  TITLE     Peptide sequences specific for the hepatic stages of P. falciparum
            bearing epitopes capable of stimulating the T lymphocytes
  JOURNAL   Patent: US 6319502-A 46 20-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .12
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 8.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
A36617
LOCUS       A36617                    23 bp    DNA     linear   PAT 05-MAR-1997
DEFINITION  Sequence 26 from Patent WO9325707.
ACCESSION   A36617
VERSION     A36617.1  GI:2293921
KEYWORDS    .
SOURCE      Human herpesvirus 3 (HHV-3)
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 23)
  AUTHORS   Tenorio,M.A.
  TITLE     METHODS FOR AMPLIFICATION OF GENOME AND MIXTURES OF INITIATOR
            OLIGONUCLEOTIDES FOR THE DETECTION AND IDENTIFICATION OF RELATED
            GENOMIC SEQUENCES
  JOURNAL   Patent: WO 9325707-A 26 23-DEC-1993;
            INST DE SALUD CARLOS III (ES)
COMMENT     Other publication AU 4327593 940104
            Other publication ES 2048652 940316.
FEATURES             Location/Qualifiers
     source          1. .23
                     /organism="Human herpesvirus 3"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10335"
ORIGIN

  Query Match           100.0%;  Score 12;  DB 6;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 6
G32838
LOCUS       G32838                   104 bp    DNA     linear   STS 24-SEP-1999
DEFINITION  A009U15 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G32838
VERSION     G32838.1  GI:5923359
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 104)
  AUTHORS   Adams,M.D.
  TITLE     Human STS sequences
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Email: mdadams@tigr.org
            Primer A: CAGCTTTCTACTTTGACAATGTA
            Primer B: ACAGGAGGTTATAGACAACAGTT
            STS size: 104
            PCR Profile:
                      Denaturation:  96C 5min
                      Anneal:        54C 30sec
                      Extend:        72C 30sec
                      Denature:      95C 30sec
                      FinalExtend:   72C  5min
                      Cycles:        30
```

```
           Protocol:
                    GenomicDNA:   25 ng
                    Primer:        0.43 uM each
                    dNTPs:       230 uM each
                    AmpliTaq:      0.5 units
                    TaqStart Ab:   0.5 units
                    Total Volume: 10 ul

           Buffer:
                    Tris-HCl pH8.8: 100 mM
                    KCl:            500 mM
                    MgCl2:           20 mM
                    Triton X-100:    1%
                    Concentration:  10X

           Prepared with primer pairs derived from THC131873: GenBank
           Accession Numbers-  R85291, T97765, L75280.
FEATURES             Location/Qualifiers
     source          1. .104
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human"
     STS             1. .104
     primer_bind     1. .23
     primer_bind     complement(82. .104)
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 104;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         62 AAGGTTATATAT 73


RESULT 7
CR378386/c
LOCUS       CR378386                113 bp    DNA     linear   STS 24-MAR-2004
DEFINITION  Arabidopsis thaliana transposon insertion STS GT_3.5710, sequence
            tagged site.
ACCESSION   CR378386
VERSION     CR378386.1  GI:45724825
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B., Walsh,S.,
            Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2004) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N174445.
FEATURES             Location/Qualifiers
     source          1. .113
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="DSG6 x AC4"
                     /db_xref="taxon:3702"
                     /clone="AL079349"
                     /ecotype="Landsberg erecta NASC stock code NW20"
     STS             1. .113
                     /standard_name="GT_3.5710"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         15 AAGGTTATATAT 4


RESULT 8
BX294638
LOCUS       BX294638                139 bp    DNA     linear   STS 11-JUN-2003
DEFINITION  Arabidopsis thaliana transposon insertion STS SM_3.3404, sequence
            tagged site.
ACCESSION   BX294638
VERSION     BX294638.1  GI:29125649
KEYWORDS    STS; STS, sequence tagged site.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B.,
            Murphy,G., Langham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139)
  AUTHORS   Clarke,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Clarke J.H., John Innes Centre, Colney
            Lane, Norwich, NR4 7UJ, UK
COMMENT     AT denotes an activation tag dissociation transposon within a
            single line, ET an enhancer trap dissociation transposon, GT a gene
            trap dissociation transposon, MT a mis-expression enhancer trap
            dissociation transposon, SM a defective suppressor mutator
            transposon. _3 denotes a sequence derived from the 3'end of the
            transposon, _5 denotes a sequence derived from the 5'end of the
            transposon BBSRC GARNet, ATIS project
            On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
            code: N57315.
FEATURES             Location/Qualifiers
     source          1. .139
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia-0 NASC stock code N1092"
                     /db_xref="taxon:3702"
                     /clone="AC002062"
                     /note="Derived from superpool NASC code Unknown"
     STS             1. .139
                     /standard_name="SM_3.3404"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 139;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         95 AAGGTTATATAT 106


RESULT 9
BD043274/c
LOCUS       BD043274                146 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD043274
```

```
VERSION     BD043274.1  GI:22585016
KEYWORDS    JP 2001269182-A/19520.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 19520 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/19520
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 10
AX907741/c
LOCUS       AX907741                146 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 23604 from Patent EP1033401.
ACCESSION   AX907741
VERSION     AX907741.1  GI:40063821
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE     Expressed sequence tags and encoded human proteins
  JOURNAL   Patent: EP 1033401-A 23604 06-SEP-2000;
            Genset (FR)
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 11
AF487987/c
LOCUS       AF487987                169 bp    DNA     linear   INV 09-OCT-2002
DEFINITION  Triatoma dimidiata clone TDMS16, microsatellite sequence.
ACCESSION   AF487987
VERSION     AF487987.1  GI:19744817
KEYWORDS    .
SOURCE      Triatoma dimidiata (kissing bug)
  ORGANISM  Triatoma dimidiata
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Heteroptera;
            Panheteroptera; Cimicomorpha; Reduviidae; Triatominae; Triatoma.
REFERENCE   1  (bases 1 to 169)
  AUTHORS   Anderson,J.M., Lai,J.E., Dotson,E.M., Cordon-Rosales,C., Ponce,C.,
            Norris,D.E. and Beard,C.B.
  TITLE     Identification and characterization of microsatellite markers in
            the Chagas disease vector Triatoma dimidiata
  JOURNAL   Infect. Genet. Evol. 1 (3), 243-248 (2002)
REFERENCE   2  (bases 1 to 169)
  AUTHORS   Anderson,J.M., Lai,J.E., Dotson,E.M. and Beard,C.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) NCID/DPD/Entomology Branch, Centers for
            Disease Control and Prevention, 4770 Buford Hwy, Mail Stop F-22,
            Atlanta, GA 30341-3724, USA
FEATURES             Location/Qualifiers
     source          1. .169
                     /organism="Triatoma dimidiata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:72491"
                     /clone="TDMS16"
     repeat_region   1. .169
                     /note="microsatellite"
                     /rpt_type=tandem
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 169;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        129 AAGGTTATATAT 118


RESULT 12
BD268181/c
LOCUS       BD268181                184 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Gene to be the target for herbicide and method for targeting the
            gene.
ACCESSION   BD268181
VERSION     BD268181.1  GI:33077949
KEYWORDS    JP 2002534128-A/8.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Levin,J.Z., Budziszewski,G.J., Louis,S.L.P. and Glover,L.W.
  TITLE     Gene to be the target for herbicide and method for targeting the
  JOURNAL   Patent: JP 2002534128-A 8 15-OCT-2002;
            SYNGENTA PARTICIPATIONS AG
COMMENT     OS   Arabidopsis thaliana (thale cress)
            PN   JP 2002534128-A/8
            PD   15-OCT-2002
            PF   13-JAN-2000 JP 2000593762
            PR   15-JAN-1999 US    09/232760,26-JAN-1999 US    09/237479 PR
            03-FEB-1999 US    09/244288,18-FEB-1999 US    09/252336 PR
            30-MAR-1999 US    09/281376
            PI   JOSHUA ZVI LEVIN,GREGORY JOSEPH BUDZISZEWSKI,SHARON LEE POTTER

            PI   LOUIS,
```

```
            PI   LYN WEGRICH GLOVER
            PC   C12N15/09,A01H5/00,C12N1/19,C12N1/21,C12N5/10,C12Q1/02 PC
            ,G01N33/15,G01N33/50,
            PC   G01N33/566,C12N15/00,C12N5/00
            CC   Gene to be the target for herbicide and method for targeting
            CC         the gene
            FH   Key             Location/Qualifiers
            FT   source          1. .184
            FT                   /organism='Arabidopsis thaliana (thale  FT
                       cress)'.
FEATURES             Location/Qualifiers
     source          1. .184
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 13
AR210010/c
LOCUS       AR210010                184 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 13 from patent US 6387637.
ACCESSION   AR210010
VERSION     AR210010.1  GI:21512129
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Levin,J.Z., Budziszewski,G.J., Potter,S.L. and Wegrich,L.M.
  TITLE     Herbicide target genes and method
  JOURNAL   Patent: US 6387637-A 13 14-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .184
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 14
AX196591
LOCUS       AX196591                188 bp    DNA     linear   PAT 07-SEP-2001
DEFINITION  Sequence 298 from Patent WO0151627.
ACCESSION   AX196591
VERSION     AX196591.1  GI:15386797
KEYWORDS    .
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1
  AUTHORS   Hauge,B.M., Wang,M.L., Parsons,J.D. and Parnell,L.D.
  TITLE     Nucleic acid molecules and other molecules associated with soybean
            cyst nematode resistance
  JOURNAL   Patent: WO 0151627-A 298 19-JUL-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .188
                     /organism="Glycine max"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3847"
                     /note="Seq ID: 318013_region_A3__153114_12"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 15
G23922
LOCUS       G23922                  190 bp    DNA     linear   STS 31-MAY-1996
DEFINITION  human STS WI-16913, sequence tagged site.
ACCESSION   G23922
VERSION     G23922.1  GI:1344248
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped STSs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: TACTGGGCTTGCAAAAACCC
            Primer B: GACTTTTGCTGTGCCCCAGT
            STS size: 150
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Derived from dbEST (genbank accession T97765).
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
                /map="338.9 cR from top of Chr17 linkage group"
     STS        1. .150
     primer_bind 1. .20
     primer_bind complement(131. .150)
ORIGIN

  Query Match          100.0%;  Score 12;  DB 10;  Length 190;
  Best Local Similarity  100.0%;  Pred. No. 4e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         85 AAGGTTATATAT 96


RESULT 16
AB153981/c
LOCUS       AB153981                200 bp    DNA     linear   STS 02-APR-2005
DEFINITION  Homo sapiens DNA, STS on chromosome 11, D11S0972i, sequence tagged
            site.
ACCESSION   AB153981
VERSION     AB153981.1  GI:62174799
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tamiya,G., Shinya,M., Ikuta,T., Makino,S., Matsumoto,T., Mano,S.,
            Furugaki,K., Ando,S., Nozaki,Y., Yukawa,W., Nakashige,R.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Nakami,Y., Takayama,S.,
            Endo,T., Saruwatari,T., Yagura,M., Yoshikawa,Y., Fujimoto,K.,
            Oka,A., Chiku,S., Linsen,S.E., Giphart,M.J., Bahram,S.,
            Kulski,Y.J., Fukazawa,T., Hashimoto,H., Hoshina,Y., Suzuki,Y.,
            Hotta,T., Mochida,J., Minezaki,T., Komai,K., Shiozawa,S.,
            Taniguchi,A., Yamanaka,H., Kamatani,N., Imanishi,T., Gojobori,T.
            and Inoko,H.
  TITLE     A Whole Genome Association Study of Rheumatoid Arthritis using
            27,039 Microsatellite Makers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200)
  AUTHORS   Tamiya,G., Makino,S., Fujimoto,K., Oka,A., Hayashi,H., Denda,A.,
            Linsen,S.E., Ikuta,T., Shinya,M., Endo,T., Tomizawa,M., Tokubo,E.,
            Sato,R., Takaki,A., Nagatsuka,Y., Watanabe,H., Adachi,S.,
            Makino,Y., Nakano,S., Yamamoto,A., Yoshida,K., Okamoto,K.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Takayama,S., Nakami,Y.,
            Saruwatari,T., Brand,A., van Hilten,J.A., van de Watering,L.M.,
            Giphart,M.J., Bahram,S., Kulski,Y.J. and Inoko,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Hidetoshi Inoko, Tokai University School of
            Medicine, Department of Genetic Information; Bohseidai, Isehara,
            Kanagawa, 259-1193, Japan (E-mail:hinoko@is.icc.u-tokai.ac.jp,
            Tel:81-463-93-1121, Fax:81-463-94-8884)
COMMENT     Polymorphisms were confirmed by comparing with the pooled DNA
            typing data of 88 Dutch population or 100 Australian. When there
            was no significant difference between the Japanese and these
            Caucasians, individual typing was performed to confirm
            polymorphisms.
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q14.3"
                     /note="pooled DNA of 100 Japanese unrelated individuals
                     sequence tagged site D11S0972i"
     primer_bind     1. .23
                     /note="5' primer: GGATAAAGTAATAGTGCTTGCAA"
                     /PCR_conditions="denaturation 96degC 5 min, 57degC 1 min,
                     72degC 1 min; 40 cycles 96degC 45 sec, 57degC 45 sec,
                     72degC 1 min"
     repeat_region   67. .177
                     /note="sequence tagged site D11S0972i"
                     /rpt_type=tandem
                     /rpt_unit="ac"
                     /rpt_unit="at"
                     /rpt_unit="ta"
     primer_bind     complement(177. .200)
                     /note="3' primer: CATAGCTACTCTCACATGAAGGTT"
ORIGIN

  Query Match          100.0%;  Score 12;  DB 10;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        182 AAGGTTATATAT 171


RESULT 17
ATH531582
LOCUS       ATH531582               201 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            237F12.
ACCESSION   AJ531582
VERSION     AJ531582.1  GI:26799842
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="237F12"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .201
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match          100.0%;  Score 12;  DB 15;  Length 201;
  Best Local Similarity  100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        116 AAGGTTATATAT 127


RESULT 18
ATH531598
LOCUS       ATH531598               201 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, left border, clone
            238A11.
ACCESSION   AJ531598
VERSION     AJ531598.1  GI:26799858
KEYWORDS    left border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 201)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="238A11"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .201
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        116 AAGGTTATATAT 127


RESULT 19
G31833
LOCUS       G31833                  225 bp    DNA     linear   STS 28-SEP-1998
DEFINITION  SpAGH1 Norway spruce endosperms Picea abies STS genomic, sequence
            tagged site.
ACCESSION   G31833
VERSION     G31833.1  GI:1927164
KEYWORDS    STS.
SOURCE      Picea abies (Norway spruce)
  ORGANISM  Picea abies
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Pfeiffer,A., Olivieri,A.M. and Morgante,M.
  TITLE     Identification and characterization of microsatellites in Norway
            spruce (Picea abies K.)
  JOURNAL   Genome 40 (4), 411-419 (1997)
   PUBMED   9276931
COMMENT
            Contact: Michele Morgante
            Dipartimento di Produzione Vegetale e Tecnologie Agrarie
            Universita' degli Studi di Udine
            Via delle Scienze 208, I-33100 Udine, ITALY
            Tel: 39432558606
            Fax: 39432558603
            Email: morgante@dpvta.uniud.it
            Primer A: Unknown
            Primer B: Unknown.
FEATURES             Location/Qualifiers
     source          1. .225
                     /organism="Picea abies"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3329"
                     /tissue_type="Endosperm"
                     /clone_lib="Norway spruce endosperms"
                     /note="Vector: lambda ZAP II; V-type: Phage; Size selected
                     (200-600 bp) genomic fragments obtained through
                     restriction enzyme digestion cloned into lambda ZapII"
     STS             <1. .>225
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        117 AAGGTTATATAT 128


RESULT 20
AY020751/c
LOCUS       AY020751                226 bp    DNA     linear   PLN 07-FEB-2001
DEFINITION  Oryza sativa microsatellite MRG3076 containing (TA)X13, genomic
            sequence.
ACCESSION   AY020751
VERSION     AY020751.1  GI:12703793
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 226)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Simple sequence repeats from Monsanto rice genomic sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
            Blvd., Creve Coeur, MO 63167, USA
COMMENT     Derived from rice genomic sequences generated from the Monsanto
            Rice Genome Sequencing project. Please see
            http://www.rice-research.org for more information.  The sequence
            data were produced primarily in the laboratories of Dr. Leroy Hood
            at the University of Washington in Seattle.
FEATURES             Location/Qualifiers
     source          1. .226
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
```

```
     repeat_region   1. .226
                     /note="microsatellite MRG3076"
                     /rpt_type=tandem
                     /rpt_unit="ta"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        131 AAGGTTATATAT 120


RESULT 21
MMU409475
LOCUS       MMU409475                227 bp    mRNA    linear   ROD 27-JUN-2001
DEFINITION  Mus musculus RNA binding site for Dazl protein, clone cg10.
ACCESSION   AJ409475
VERSION     AJ409475.1  GI:13016754
KEYWORDS    Dazl gene; RNA binding protein.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 227)
  AUTHORS   Venables,J.P., Ruggiu,M. and Cooke,H.J.
  TITLE     The RNA-binding specificity of the mouse Dazl protein
  JOURNAL   Nucleic Acids Res. 29 (12), 2479-2483 (2001)
   PUBMED   11410654
REFERENCE   2  (bases 1 to 227)
  AUTHORS   Venables,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2001) Venables J.P., Chromosome Biology, MRC
            Human Gentics Unit, Western Genral Hospital, Crewe Rd, Edinburgh
            EH4 2XU, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .227
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="cg10"
     protein_bind    1. .227
                     /note="U-rich, Dazl binding site"
                     /bound_moiety="RNA binding protein, Dazl"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 9;  Length 227;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        207 AAGGTTATATAT 218


RESULT 22
AB143702/c
LOCUS       AB143702                 232 bp    DNA     linear   STS 02-APR-2005
DEFINITION  Homo sapiens DNA, STS on chromosome 20, D20S0259i, sequence tagged
            site.
ACCESSION   AB143702
VERSION     AB143702.1  GI:62164520
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tamiya,G., Shinya,M., Ikuta,T., Makino,S., Matsumoto,T., Mano,S.,
            Furugaki,K., Ando,S., Nozaki,Y., Yukawa,W., Nakashige,R.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Nakami,Y., Takayama,S.,
            Endo,T., Saruwatari,T., Yagura,M., Yoshikawa,Y., Fujimoto,K.,
            Oka,A., Chiku,S., Linsen,S.E., Giphart,M.J., Bahram,S.,
            Kulski,Y.J., Fukazawa,T., Hashimoto,H., Hoshina,Y., Suzuki,Y.,
            Hotta,T., Mochida,J., Minezaki,T., Komai,K., Shiozawa,S.,
            Taniguchi,A., Yamanaka,H., Kamatani,N., Imanishi,T., Gojobori,T.
            and Inoko,H.
  TITLE     A Whole Genome Association Study of Rheumatoid Arthritis using
            27,039 Microsatellite Makers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232)
  AUTHORS   Tamiya,G., Makino,S., Fujimoto,K., Oka,A., Hayashi,H., Denda,A.,
            Linsen,S.E., Ikuta,T., Shinya,M., Endo,T., Tomizawa,M., Tokubo,E.,
            Sato,R., Takaki,A., Nagatsuka,Y., Watanabe,H., Adachi,S.,
            Makino,Y., Nakano,S., Yamamoto,A., Yoshida,K., Okamoto,K.,
            Yamaguchi,D., Ishibashi,H., Yonekura,M., Takayama,S., Nakami,Y.,
            Saruwatari,T., Brand,A., van Hilten,J.A., van de Watering,L.M.,
            Giphart,M.J., Bahram,S., Kulski,Y.J. and Inoko,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Hidetoshi Inoko, Tokai University School of
            Medicine, Department of Genetic Information; Bohseidai, Isehara,
            Kanagawa, 259-1193, Japan (E-mail:hinoko@is.icc.u-tokai.ac.jp,
            Tel:81-463-93-1121, Fax:81-463-94-8884)
COMMENT     Polymorphisms were confirmed by comparing with the pooled DNA
            typing data of 88 Dutch population or 100 Australian. When there
            was no significant difference between the Japanese and these
            Caucasians, individual typing was performed to confirm
            polymorphisms.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20p12.3"
                     /note="pooled DNA of 100 Japanese unrelated individuals
                     sequence tagged site D20S0259i"
     primer_bind     1. .24
                     /note="5' primer: CAAACATTGAACCATTACTTCTAG"
                     /PCR_conditions="denaturation 96degC 5 min, 57degC 1 min,
                     72degC 1 min; 40 cycles 96degC 45 sec, 57degC 45 sec,
                     72degC 1 min"
     repeat_region   128. .202
                     /note="sequence tagged site D20S0259i"
                     /rpt_type=tandem
                     /rpt_unit="ta"
                     /rpt_unit="tatatg"
     primer_bind     complement(210. .232)
                     /note="3' primer: GGTCATGAGTTCAAGGTTAATAC"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 85


RESULT 23
AY020956
LOCUS       AY020956                 232 bp    DNA     linear   PLN 07-FEB-2001
DEFINITION  Oryza sativa microsatellite MRG3281 containing (TA)X16, genomic
            sequence.
ACCESSION   AY020956
VERSION     AY020956.1  GI:12703998
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Simple sequence repeats from Monsanto rice genomic sequences
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232)
  AUTHORS   Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
            Blvd., Creve Coeur, MO 63167, USA
COMMENT     Derived from rice genomic sequences generated from the Monsanto
            Rice Genome Sequencing project. Please see
            http://www.rice-research.org for more information.  The sequence
            data were produced primarily in the laboratories of Dr. Leroy Hood
            at the University of Washington in Seattle.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
     repeat_region   1. .232
                     /note="microsatellite MRG3281"
                     /rpt_type=tandem
                     /rpt_unit="ta"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 15;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 107


RESULT 24
AY114172
LOCUS       AY114172                234 bp    DNA     linear   VRL 26-FEB-2003
DEFINITION  Human herpesvirus 3 DNA polymerase gene, partial cds.
ACCESSION   AY114172
VERSION     AY114172.2  GI:28566755
KEYWORDS    .
SOURCE      Human herpesvirus 3 (HHV-3)
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Detection and typing of human herpesviruses 1-5 by consensus primer
            PCR and RFLP
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2002) Pediatric, Chulalongkorn University, Rama
            IV, Patumwan, Bangkok 10330, Thailand
REFERENCE   3  (bases 1 to 234)
  AUTHORS   Sakulwira,K., Vanapongtipagorn,P., Theamboonlers,A. and
            Poovorawan,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2003) Pediatric, Chulalongkorn University, Rama
            IV, Patumwan, Bangkok 10330, Thailand
  REMARK    Sequence update by submitter
COMMENT     On Feb 26, 2003 this sequence version replaced gi:21724220.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Human herpesvirus 3"
                     /proviral
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10335"
                     /note="VZV"
     CDS             <1. .>234
                     /codon_start=1
                     /product="DNA polymerase"
                     /protein_id="AAM75496.2"
                     /db_xref="GI:28566756"
                     /translation="CNSVYGFTGVAQGFLPCLYVAATVTTIGRQMLLSTRDYIHNNWA
                     AFERFITAFPDIESSVLSQKAYEVKVIYGDTDSV"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 13;  Length 234;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        205 AAGGTTATATAT 216


RESULT 25
G37836/c
LOCUS       G37836                  236 bp    DNA     linear   STS 24-MAR-1999
DEFINITION  wh_hsp1 Plasmodium falciparum haploid Plasmodium falciparum STS
            genomic, sequence tagged site.
ACCESSION   G37836
VERSION     G37836.1  GI:3015442
KEYWORDS    STS.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Su,X.-z., Ferdig,M.T. and Wellems,T.E.
  TITLE     Integrated genetic and physical maps of Plasmodium falciparum
  JOURNAL   Unpublished (1998)
COMMENT     Synonyms: wh_hsp1
            Contact: Thomas E. Wellems
            LPD
            NIAID
            Bldg. 4, Room 126, NIH Campus, Bethesda, MD 20892-0425
            Tel: (301) 496-4021
            Fax: (301) 402-0079
            Email: tew@helix.nih.gov
            Primer A: GCATATTATTATGAATTACC
            Primer B: ATCAAAAGGTTATATATGTA
            STS size: 236
            PCR Profile:
            Initial heat: 2min at 94oC
            Cycles:       30 cycles, each having steps 20sec at 94oC;
                          10sec at 45oC; 10sec at 40oC; 30sec at 60oC
            Machine:      Perkin Elmer GeneAmp 9600
            Protocol:
                  Template:        1-10 ng
                  Primer:          5 pM each
                  dNTPs:           0.3 ul of 10 mM stock solution
                  Taq Polymerase:  0.5 unit
                  Total Volume:    15 ul

            Buffer:
                  10 mM Tris HCl, pH 8.3
                  1.5 mM MgCl2
                  50 mM KCl.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3xDd2"
                     /db_xref="taxon:5833"
                     /map="9"
                     /clone_lib="Plasmodium falciparum haploid"
                     /dev_stage="erythrocytic haploid stage"
     STS             1. .236
     primer_bind     1. .20
     primer_bind     complement(217. .236)
ORIGIN
```

```
  Query Match             100.0%;  Score 12;  DB 10;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        231 AAGGTTATATAT 220


RESULT 26
G44501/c
LOCUS       G44501                   236 bp    DNA     linear   STS 24-MAR-1999
DEFINITION  C9M106 Plasmodium falciparum haploid Plasmodium falciparum STS
            genomic, sequence tagged site.
ACCESSION   G44501
VERSION     G44501.2  GI:4511820
KEYWORDS    STS.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Su,X.-z., Ferdig,M.T. and Wellems,T.E.
  TITLE     Integrated genetic and physical maps of Plasmodium falciparum
  JOURNAL   Unpublished (1998)
COMMENT     On Mar 25, 1999 this sequence version replaced gi:4218688.

            Contact: Thomas E. Wellems
            LPD
            NIAID
            Bldg. 4, Room 126, NIH Campus, Bethesda, MD 20892-0425
            Tel: (301) 496-4021
            Fax: (301) 402-0079
            Email: tew@helix.nih.gov
            Primer A: CATATTTATTTATTCATGTCC
            Primer B: ATCAAAAGGTTATATATGTA
            STS size: 236
            PCR Profile:
            Initial heat: 2min at 94oC
            Cycles:       30 cycles, each having steps 20sec at 94oC;
                          10sec at 45oC; 10sec at 40oC; 30sec at 60oC
            Machine:      Perkin Elmer GeneAmp 9600
            Protocol:
                  Template:        1-10 ng
                  Primer:          5 pM each
                  dNTPs:           0.3 ul of 10 mM stock solution
                  Taq Polymerase:  0.5 unit
                  Total Volume:    15 ul

            Buffer:
                   10 mM Tris HCl, pH 8.3
                   1.5 mM MgCl2
                   50 mM KCl.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3xDd2"
                     /db_xref="taxon:5833"
                     /map="9"
                     /clone_lib="Plasmodium falciparum haploid"
                     /dev_stage="erythrocytic haploid stage"
     STS             1. .236
     primer_bind     1. .21
     primer_bind     complement(217. .236)
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        231 AAGGTTATATAT 220


RESULT 27
AU026951/c
LOCUS       AU026951                 238 bp    DNA     linear   STS 02-MAR-1999
DEFINITION  Rattus norvegicus, OTSUKA clone, OT62.26/819g05, microsatellite
            sequence, sequence tagged site.
ACCESSION   AU026951
VERSION     AU026951.1  GI:4516874
KEYWORDS    STS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A.,
            Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M.,
            Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,T., Nakamura,Y. and
            Tanigami,A.
  TITLE     The large-scale mapping of rat microsatellite markers
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238)
  AUTHORS   Watanabe,T.K.
  TITLE     Direct Submission                                             .
  JOURNAL   Submitted (24-JUL-1998) Takeshi K Watanabe, Otsuka Pharmaceutical
            Co., Ltd., Otsuka GEN Research Institute; 463-10, Kagasuno,
            Kawauchi-cho, Tokushima, Tokushima 771-0192, Japan
            (E-mail:watanabe@otsuka.genome.ad.jp, Tel:81-886-65-2888,
            Fax:81-886-37-1035)
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="OT62.26/819g05"
                     /note="OT62.26/819g05F=5'-ATGCGACAGAAGCCAGAAAC-3',
                     OT62.26/819g05R=5'-CAGTTAGCCTTGAGCTTAAAGCA-3'"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        209 AAGGTTATATAT 198


RESULT 28
BD028743/c
LOCUS       BD028743                 242 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD028743
VERSION     BD028743.1  GI:22570485
KEYWORDS    JP 2001269182-A/4989.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 4989 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/4989
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
```

```
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers.
FEATURES              Location/Qualifiers
     source           1. .242
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 29
AX893210/c
LOCUS       AX893210                 242 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 9073 from Patent EP1033401.
ACCESSION   AX893210
VERSION     AX893210.1  GI:40048094
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE     Expressed sequence tags and encoded human proteins
  JOURNAL   Patent: EP 1033401-A 9073 06-SEP-2000;
            Genset (FR)
FEATURES              Location/Qualifiers
     source           1. .242
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 30
CQ457358
LOCUS       CQ457358                 255 bp    DNA     linear   PAT 30-JAN-2004
DEFINITION  Sequence 136 from Patent WO0192581.
ACCESSION   CQ457358
VERSION     CQ457358.1  GI:41422977
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Algate,P.A., Harlocker,S.L. and Jones,R.
  TITLE     Compositions and methods for the therapy and diagnosis of
            ovariancancer
  JOURNAL   Patent: WO 0192581-A 136 06-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES              Location/Qualifiers
     source           1. .255
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37


Search completed: December 20, 2005, 10:43:13
Job time : 70.5382 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 04:22:54 ; Search time 9.22892 Seconds
(without alignments)
8665.841 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	13	5	ABF93204	Abf93204 Oligonucl
c	2	12	100.0	13	5	ABF93205	Abf93205 Oligonucl
	3	12	100.0	23	2	AAQ53332	Aaq53332 Positive
	4	12	100.0	23	2	AAT37524	Aat37524 Varicella
	5	12	100.0	100	12	ACH88864	Ach88864 Human gen
	6	12	100.0	121	11	ADZ42781	Adz42781 Human gen
c	7	12	100.0	146	3	AAC19529	Aac19529 Human sec
	8	12	100.0	146	12	ACH81646	Ach81646 Human gen
c	9	12	100.0	184	3	AAA50085	Aaa50085 Arabidops
	10	12	100.0	188	5	AAI61667	Aai61667 Soybean 3
c	11	12	100.0	242	3	AAC04998	Aac04998 Human sec
	12	12	100.0	255	6	ABL77158	Abl77158 Human ova
	13	12	100.0	267	6	ABN77348	Abn77348 Human ORF
c	14	12	100.0	277	5	ABV07118	Abv07118 Human pro
c	15	12	100.0	289	6	ABL77086	Abl77086 Human ova
c	16	12	100.0	294	3	AAC18502	Aac18502 Human sec
c	17	12	100.0	313	6	ABL80474	Abl80474 Human ova
	18	12	100.0	316	6	ABK46190	Abk46190 cDNA enco
	19	12	100.0	317	4	AAH69998	Aah69998 Human cer
	20	12	100.0	317	4	AAH73195	Aah73195 Human cer
	21	12	100.0	320	4	AAH70509	Aah70509 Human cer
	22	12	100.0	341	4	AAH69769	Aah69769 Human cer
	23	12	100.0	362	3	AAC24388	Aac24388 Human sec
	24	12	100.0	366	4	AAH71423	Aah71423 Human cer
	25	12	100.0	367	4	AAH35822	Aah35822 Human col
	26	12	100.0	374	4	AAS37872	Aas37872 Novel hum
c	27	12	100.0	375	13	ADV83971	Adv83971 Streptoco
	28	12	100.0	378	4	AAK81895	Aak81895 Human imm
	29	12	100.0	378	4	AAK81894	Aak81894 Human imm
c	30	12	100.0	378	4	AAK62195	Aak62195 Human imm
	31	12	100.0	385	10	ABT22768	Abt22768 Breast ca
c	32	12	100.0	402	4	AAK87775	Aak87775 Human dig
c	33	12	100.0	402	5	AAS31692	Aas31692 Human liv
c	34	12	100.0	402	6	ABN90047	Abn90047 Human liv
c	35	12	100.0	402	11	ADJ14814	Adj14814 Human liv
c	36	12	100.0	405	4	AAK60591	Aak60591 Human imm
	37	12	100.0	405	9	ACH18939	Ach18939 Human adu
	38	12	100.0	410	8	ABX44501	Abx44501 Bovine ES
c	39	12	100.0	411	10	ADI21471	Adi21471 Novel hum
	40	12	100.0	412	6	ABK64663	Abk64663 Human ben
	41	12	100.0	420	14	AEB90163	Aeb90163 Isolated
c	42	12	100.0	421	8	ABZ52835	Abz52835 Aspergill
c	43	12	100.0	422	4	AAI87578	Aai87578 Human pol
c	44	12	100.0	423	13	ACF85182	Acf85182 Human SIR
c	45	12	100.0	432	5	ABV37059	Abv37059 Human pro

ALIGNMENTS

RESULT 1
ABF93204
ID ABF93204 standard; DNA; 13 BP.
XX
AC ABF93204;
XX
DT 22-FEB-2002 (first entry)
XX
DE Oligonucleotide SEQ ID NO 193201 for detecting SNP TSC0047525.
XX
KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS Homo sapiens.
XX
PN WO200177384-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-IB000713.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-657177/75.
XX
PT Set of oligonucleotides, useful for diagnosis and cell typing, is
PT designed to detect single-nucleotide polymorphisms and cytosine
PT methylation status.
XX
PS Claim 1; SEQ ID NO 193201; 29pp + Sequence Listing; German.
XX
CC This invention describes novel oligonucleotide primers or peptide nucleic
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC and cytosine methylation status in chemically pretreated genomic DNA. The
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC range of diseases including immune system, gastrointestinal, respiratory,
CC central nervous system, cardiovascular and metabolic disorders. The

```
CC   oligomers are also used for detecting cell type differentiation. ABC00010
CC   -ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and ABI00010-ABI82073
CC   represent the oligomers described in the invention. NOTE: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13 BP; 5 A; 0 C; 2 G; 5 T; 0 U; 1 Other;

  Query Match            100.0%;  Score 12;  DB 5;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 5.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
ABF93205/c
ID   ABF93205 standard; DNA; 13 BP.
XX
AC   ABF93205;
XX
DT   22-FEB-2002  (first entry)
XX
DE   Oligonucleotide SEQ ID NO 193202 for detecting SNP TSC0047525.
XX
KW   SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW   peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW   central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS   Homo sapiens.
XX
PN   WO200177384-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-IB000713.
XX
PR   07-APR-2000; 2000DE-01019173.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-657177/75.
XX
PT   Set of oligonucleotides, useful for diagnosis and cell typing, is
PT   designed to detect single-nucleotide polymorphisms and cytosine
PT   methylation status.
XX
PS   Claim 1; SEQ ID NO 193202; 29pp + Sequence Listing; German.
XX
CC   This invention describes novel oligonucleotide primers or peptide nucleic
CC   acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC   and cytosine methylation status in chemically pretreated genomic DNA. The
CC   oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC   range of diseases including immune system, gastrointestinal, respiratory,
CC   central nervous system, cardiovascular and metabolic disorders. The
CC   oligomers are also used for detecting cell type differentiation. ABC00010
CC   -ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and ABI00010-ABI82073
CC   represent the oligomers described in the invention. NOTE: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 13 BP; 5 A; 2 C; 0 G; 5 T; 0 U; 1 Other;

  Query Match            100.0%;  Score 12;  DB 5;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 5.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         13 AAGGTTATATAT 2


RESULT 3
AAQ53332
ID   AAQ53332 standard; DNA; 23 BP.
XX
AC   AAQ53332;
XX
DT   25-MAR-2003  (revised)
DT   30-JUN-1994  (first entry)
XX
DE   Positive primer for detection of VZV.
XX
KW   Varicella Zoster Virus; Herpes viridae; amplification reaction;
KW   DNA polymerase; conserved sequence; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .6
FT                   /*tag=  a
FT                   /note= "correponds to non-conserved region of herpes
FT                   viridae DNA polymerase"
FT   misc_feature    7. .23
FT                   /*tag=  b
FT                   /note= "corresponds to conserved region of herpes viridae
FT                   DNA polymerase"
XX
PN   WO9325707-A2.
XX
PD   23-DEC-1993.
XX
PF   04-JUN-1993;   93WO-ES000048.
XX
PR   05-JUN-1992;   92ES-00001174.
XX
PA   (SALU-) INST SALUD CARLOS III.
XX
PI   Tenorio Matanzo A;
XX
DR   WPI; 1994-007564/01.
XX
PT   Amplification of genome(s) and initiator oligo-nucleotide(s) mixts. -
PT   using single reaction mixt. to detect and identify infections by related
PT   viruses.
XX
PS   Claim 14(i); Page 26; 41pp; Spanish.
XX
CC   The known amino acid sequences of DNA polymerases from 6 different
CC   members of the herpes viridae family were aligned so that conserved
CC   regions could be identified. A set of primers (AAQ53331-5) was designed
CC   based on the conserved amino acid motif IYGDTD and having 6 non-conserved
CC   nucleotides at the 5'-end. Sequence AAQ53332 is specific for VZV.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 23 BP; 9 A; 1 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 12;  DB 2;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
AAT37524
ID   AAT37524 standard; DNA; 23 BP.
XX
```

```
AC   AAT37524;
XX
DT   23-APR-1997  (first entry)
XX
DE   Varicella-zoster virus genomic amplification primer.
XX
KW   Polymerase chain reaction; PCR primer; Herpesviridae; Picornaviridae;
KW   human herpesvirus; enterovirus; diagnosis; detection; ss.
XX
OS   Varicella-zoster virus.
XX
PN   WO9625909-A2.
XX
PD   29-AUG-1996.
XX
PF   16-FEB-1996;   96WO-ES000031.
XX
PR   17-FEB-1995;   95ES-00000320.
XX
PA   (SALU-) INST SALUD CARLOS III.
XX
PI   Casas I,  Tenorio A,  Echevarria JM,  Klapper PE,  Cleator GM;
XX
DR   WPI; 1996-402109/40.
XX
PT   Differential diagnosis of infectious agents by genomic amplification -
PT   using primer mixts., each specific for a family of infectious agents,
PT   esp. for distinguishing human herpes virus and entero-virus infection.
XX
PS   Claim 13; Page 29; 55pp; Spanish.
XX
CC   A mixture of amplification primers was designed which could amplify both
CC   enterovirus and herpesvirus sequences from the same sample. The mixture
CC   contained at least one of the primers in AAT37520-T37522 and at least one
CC   of the primers in AAT37523-T37532. A second PCR can be performed on the
CC   products of the first reaction in order to distinguish specific viruses.
CC   Alternatively, the PCR products can be detected by hybridisation to virus
CC   -specific probes
XX
SQ   Sequence 23 BP; 9 A; 1 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
ACH88864
ID   ACH88864 standard; DNA; 100 BP.
XX
AC   ACH88864;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #22059.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 1; SEQ ID NO 22059; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 100 BP; 31 A; 21 C; 27 G; 21 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 12;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         44 AAGGTTATATAT 55


RESULT 6
ADZ42781
ID   ADZ42781 standard; DNA; 121 BP.
XX
AC   ADZ42781;
XX
DT   16-JUN-2005  (first entry)
XX
DE   Human gene encoding a receptor protein with an A61G SNP Seq 363.
XX
KW   SNP detection; DNA amplification; DNA microarray;
```

```
KW   single nucleotide polymorphism; SNP; selectable marker; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       61
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   JP2003274966-A.
XX
PD   30-SEP-2003.
XX
PF   25-MAR-2002; 2002JP-00084465.
XX
PR   25-MAR-2002; 2002JP-00084465.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2003-869435/81.
XX
PT   New polynucleotide isolated from human gene useful for detection of
PT   single nucleotide polymorphisms.
XX
PS   Claim 1; SEQ ID NO 363; 687pp; Japanese.
XX
CC   This invention relates to novel polynucleotides isolated and purified
CC   from 99 human genes that comprises any one of 863 fully defined sequences
CC   of 121 nucleotides, as given in specification. Specifically, it refers to
CC   these polynucleotides as each containing a single base substitution
CC   (otherwise known as a single nucleotide polymorphism) that are useful as
CC   genetic markers and for diagnosing certain diseases. The present
CC   invention provides oligonucleotides containing these SNPs that encode
CC   receptor proteins, as well as PCR primers, labeling probes and a
CC   microarray. Accordingly, this array is useful for diagnosing disease and
CC   determining side effects of a medical agent. This polynucleotide sequence
CC   is a human gene containing a SNP, as given in an exemplification of the
CC   invention.
XX
SQ   Sequence 121 BP; 32 A; 17 C; 24 G; 47 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 12;  DB 11;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         33 AAGGTTATATAT 44


RESULT 7
AAC19529/c
ID   AAC19529 standard; cDNA; 146 BP.
XX
AC   AAC19529;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 23604.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 23604; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 146 BP; 38 A; 44 C; 30 G; 34 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 3;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 8
ACH81646
ID   ACH81646 standard; DNA; 146 BP.
XX
AC   ACH81646;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human genome derived single exon probe #14841.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG,  Rank DR,  Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
```

XX
PS Claim 1; SEQ ID NO 14841; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above- mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ Sequence 146 BP; 41 A; 33 C; 37 G; 35 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 12; Length 146;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 97 AAGGTTATATAT 108

RESULT 9
AAA50085/c
ID AAA50085 standard; cDNA; 184 BP.
XX
AC AAA50085;
XX
DT 25-OCT-2000 (first entry)
XX
DE Arabidopsis herbicide target gene 245 3' UTR cDNA.
XX
KW Herbicide; 245 gene; inhibitor; screening; transgenic plant;
KW herbicide resistance; crop protection; peptide release factor 2; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200042205-A2.
XX
PD 20-JUL-2000.
XX
PF 13-JAN-2000; 2000WO-EP000246.
XX
PR 15-JAN-1999; 99US-00232760.
PR 26-JAN-1999; 99US-00237479.
PR 03-FEB-1999; 99US-00244288.
PR 18-FEB-1999; 99US-00252336.
PR 30-MAR-1999; 99US-00281376.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Levin JZ, Budziszewski GJ, Potter SL, Wegrich LM;
XX
DR WPI; 2000-482835/42.
XX
PT New DNA, comprising a sequence that codes for a protein essential for
PT seedling growth, for identifying potential herbicides and developing
PT herbicide tolerant plants, tissues, seeds and cells.
XX
PS Claim 46; Page 87; 100pp; English.
XX
CC This is the 3' untranslated region (UTR) of cDNA corresponding to the
CC Arabidopsis thaliana 245 gene. The coding region (see AAA50078) of the
CC cDNA codes for a protein (see AAY95654) that is essential for normal
CC seedling growth and development, and which shows sequence similarity to
CC peptide release factor 2. The 245 cDNA was isolated from a cDNA library
CC using Arabidopsis peptide release factor DNA as probe. The 245 protein is
CC a useful herbicide target, and the invention provides screening assays
CC useful for identifying inhibitors that are potential herbicides. The
CC invention is also applied to using the 245 gene for the development of
CC herbicide tolerant plants, plant tissues, plant seeds, and plant cells
XX
SQ Sequence 184 BP; 84 A; 23 C; 25 G; 50 T; 0 U; 2 Other;

Query Match 100.0%; Score 12; DB 3; Length 184;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 86 AAGGTTATATAT 75

RESULT 10
AAI61667
ID AAI61667 standard; DNA; 188 BP.
XX
AC AAI61667;
XX
DT 16-OCT-2001 (first entry)
XX
DE Soybean 318O13 region A3 DNA, SEQ ID NO: 298.
XX
KW Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240O17 region G3; 318O13 region A3; 515O02 region G2; ds.
XX
OS Glycine max.
XX
PN WO200151627-A2.
XX
PD 19-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-US000552.
XX
PR 07-JAN-2000; 2000US-0174880P.
XX
PA (MONS) MONSANTO CO.
XX
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
DR WPI; 2001-425872/45.
XX
PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs.
XX
PS Claim 25; Page 1105; 1353pp; English.

```
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule provided
CC   in the specification
XX
SQ   Sequence 188 BP; 79 A; 20 C; 16 G; 73 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 5;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 11
AAC04998/c
ID   AAC04998 standard; cDNA; 242 BP.
XX
AC   AAC04998;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 9073.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 9073; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 242 BP; 55 A; 50 C; 25 G; 100 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 12;  DB 3;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 12
ABL77158
ID   ABL77158 standard; cDNA; 255 BP.
XX
AC   ABL77158;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:136.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 136; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 255 BP; 88 A; 53 C; 36 G; 78 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37
```

RESULT 13
ABN77348
ID ABN77348 standard; cDNA; 267 BP.
XX
AC ABN77348;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF2295 cDNA, SEQ ID NO:4589.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US017076.
XX
PR 24-MAY-2000; 2000US-0206690P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Leach MD, Shimkets RA;
XX
DR WPI; 2002-106200/14.
DR P-PSDB; ABP33322.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX
PS Claim 1; Page 1401; 2508pp; English.
XX
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX
SQ Sequence 267 BP; 67 A; 52 C; 58 G; 89 T; 0 U; 1 Other;

  Query Match   100.0%; Score 12; DB 6; Length 267;
  Best Local Similarity   100.0%; Pred. No. 4.5e+03;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         23 AAGGTTATATAT 34
```

RESULT 14
ABV07118/c
ID ABV07118 standard; cDNA; 277 BP.
XX
AC ABV07118;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 7109.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 1155; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

```
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 277 BP; 75 A; 51 C; 61 G; 88 T; 0 U; 2 Other;

  Query Match            100.0%;  Score 12;  DB 5;  Length 277;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         76 AAGGTTATATAT 65


RESULT 15
ABL77086/c
ID   ABL77086 standard; cDNA; 289 BP.
XX
AC   ABL77086;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:64.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 64; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 289 BP; 99 A; 36 C; 56 G; 97 T; 0 U; 1 Other;

  Query Match            100.0%;  Score 12;  DB 6;  Length 289;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        250 AAGGTTATATAT 239


RESULT 16
AAC18502/c
ID   AAC18502 standard; cDNA; 294 BP.
XX
AC   AAC18502;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 22577.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 22577; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 294 BP; 98 A; 49 C; 37 G; 107 T; 0 U; 3 Other;

  Query Match            100.0%;  Score 12;  DB 3;  Length 294;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        240 AAGGTTATATAT 229
```

```
RESULT 17
ABL80474/c
ID   ABL80474 standard; cDNA; 313 BP.
XX
AC   ABL80474;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human ovarian cancer related cDNA clone SEQ ID NO:3452.
XX
KW   Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192581-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US017756.
XX
PR   26-MAY-2000; 2000US-0207484P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA,  Harlocker SL,  Jones R;
XX
DR   WPI; 2002-122075/16.
XX
PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 3452; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 313 BP; 75 A; 54 C; 72 G; 106 T; 0 U; 6 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 313;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        303 AAGGTTATATAT 292


RESULT 18
ABK46190
ID   ABK46190 standard; cDNA; 316 BP.
XX
AC   ABK46190;
XX
DT   05-JUN-2002  (first entry)
XX
DE   cDNA encoding colon tumour protein, SEQ ID No 1741.
XX
KW   Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200212328-A2.
XX
PD   14-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US024218.
XX
PR   03-AUG-2000; 2000US-0223283P.
PR   28-MAR-2001; 2001US-0279763P.
PR   29-JUN-2001; 2001US-0302051P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   King GE,  Meagher MJ,  Xu J,  Secrist H;
XX
DR   WPI; 2002-241739/29.
XX
PT   New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT   for diagnosing, preventing, and treating colon cancer, and as markers for
PT   the progression of cancer.
XX
PS   Claim 1; SEQ ID NO 1741; 147pp; English.
XX
CC   The invention relates to polynucleotides encoding colon tumour proteins.
CC   The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis, prevention, and
CC   treatment of colon cancer. Polynucleotide sequences may be used as
CC   hybridisation probes or primers, and in the design and preparation of
CC   ribozyme molecules for inhibiting expression of tumour polypeptides and
CC   proteins in tumour cells. The compositions are useful for stimulating an
CC   immune response against cancer, particularly for the immunotherapy of
CC   colon cancer, and as markers for the progression of cancer. ABK44450-
CC   ABK46237 represent coding sequences of human colon tumour proteins of the
CC   invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC   data for this patent did not form part of the printed specification but
CC   was supplied by the European Patent Office
XX
SQ   Sequence 316 BP; 111 A; 56 C; 69 G; 76 T; 0 U; 4 Other;

  Query Match             100.0%;  Score 12;  DB 6;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        161 AAGGTTATATAT 172


RESULT 19
AAH69998
ID   AAH69998 standard; cDNA; 317 BP.
XX
AC   AAH69998;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1272.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
```

```
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 303; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 317 BP; 79 A; 64 C; 58 G; 89 T; 0 U; 27 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         50 AAGGTTATATAT 61


RESULT 20
AAH73195
ID   AAH73195 standard; cDNA; 317 BP.
XX
AC   AAH73195;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 4469.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 977; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 317 BP; 108 A; 51 C; 59 G; 99 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        265 AAGGTTATATAT 276


RESULT 21
AAH70509
ID   AAH70509 standard; cDNA; 320 BP.
XX
AC   AAH70509;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1783.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 386; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
```

```
SQ   Sequence 320 BP; 108 A; 51 C; 61 G; 100 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 4;  Length 320;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        268 AAGGTTATATAT 279


RESULT 22
AAH69769
ID   AAH69769 standard; cDNA; 341 BP.
XX
AC   AAH69769;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1043.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US033312.
XX
PR   08-DEC-1999;   99US-0169681P.
PR   21-DEC-1999;   99US-0171350P.
PR   14-MAR-2000; 2000US-0189315P.
PR   12-MAY-2000; 2000US-0203791P.
PR   09-JUN-2000; 2000US-0210600P.
PR   21-JUL-2000; 2000US-0220114P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer and
PT   for assessing and detecting compounds for treating the cancer.
XX
PS   Claim 1; Page 272; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy
XX
SQ   Sequence 341 BP; 110 A; 58 C; 69 G; 104 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 4;  Length 341;
  Best Local Similarity  100.0%;  Pred. No. 4.5e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        290 AAGGTTATATAT 301


RESULT 23
AAC24388
ID   AAC24388 standard; cDNA; 362 BP.
XX
AC   AAC24388;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 28463.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 28463; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 362 BP; 105 A; 74 C; 110 G; 73 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 12;  DB 3;  Length 362;
  Best Local Similarity  100.0%;  Pred. No. 4.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        295 AAGGTTATATAT 306


RESULT 24
AAH71423
ID   AAH71423 standard; cDNA; 366 BP.
XX
AC   AAH71423;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 2697.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
```

PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US033312.
XX
PR 08-DEC-1999; 99US-0169681P.
PR 21-DEC-1999; 99US-0171350P.
PR 14-MAR-2000; 2000US-0189315P.
PR 12-MAY-2000; 2000US-0203791P.
PR 09-JUN-2000; 2000US-0210600P.
PR 21-JUL-2000; 2000US-0220114P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer and
PT for assessing and detecting compounds for treating the cancer.
XX
PS Claim 1; Page 555; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy
XX
SQ Sequence 366 BP; 98 A; 85 C; 76 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 28 AAGGTTATATAT 39

RESULT 25
AAH35822
ID AAH35822 standard; cDNA; 367 BP.
XX
AC AAH35822;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:2904.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG76417.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 4589-4590; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 367 BP; 113 A; 53 C; 61 G; 133 T; 0 U; 7 Other;

Query Match 100.0%; Score 12; DB 4; Length 367;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 133 AAGGTTATATAT 144

RESULT 26
AAS37872
ID AAS37872 standard; cDNA; 374 BP.
XX
AC AAS37872;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #930.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007787.
XX
PR 09-MAR-2000; 2000US-0188609P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and treatment
PT of breast, lung and colon cancer.
XX
PS Claim 1; Page 830-831; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for

```
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention
XX
SQ   Sequence 374 BP; 127 A; 72 C; 83 G; 92 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 4;  Length 374;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        196 AAGGTTATATAT 207


RESULT 27
ADV83971/c
ID   ADV83971 standard; DNA; 375 BP.
XX
AC   ADV83971;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Streptococcus agalactiae DNA sequence, SEQ ID 5112.
XX
KW   Antibacterial; vaccine; bacterial infection; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200292818-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P,  Rusniok C,  Chevalier F,  Frangeul L,  Lalioui L;
PI   Zouine M,  Couve E,  Buchrieser C,  Poyart C,  Trieu-Cuot P,  Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
PS   Claim 4; SEQ ID NO 5112; 439pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
CC   novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC   nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC   synthesis of amino acids, cell membranes, intermediate (central)
CC   metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC   nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC   regulatory functions, replication, transcription, translation, protein
CC   transport, adaptation to atypical conditions, sensitivity to medicines
CC   and/or analogues, functions related to transposons, biosynthesis of
CC   cofactors, prosthetic groups and transporters, cell membrane proteins and
CC   cellular machinery. (I) are useful for the detection and/or amplification
CC   of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC   useful for treatment of a bacterial S. agalactiae infection. The complete
CC   genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC   present patent is an equivalent for the basic patent FR2824074A1, which
CC   contains only 2344 sequences.
XX
SQ   Sequence 375 BP; 159 A; 44 C; 64 G; 108 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 12;  DB 13;  Length 375;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        164 AAGGTTATATAT 153


RESULT 28
AAK81895
ID   AAK81895 standard; DNA; 378 BP.
XX
AC   AAK81895;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36707.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
```

PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 36707; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 378 BP; 119 A; 47 C; 76 G; 136 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 152 AAGGTTATATAT 163

RESULT 29
AAK81894
ID AAK81894 standard; DNA; 378 BP.
XX
AC AAK81894;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36706.

XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 36706; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 378 BP; 119 A; 47 C; 76 G; 136 T; 0 U; 0 Other;

Query Match 100.0%; Score 12; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 152 AAGGTTATATAT 163

RESULT 30
AAK62195/c
ID AAK62195 standard; cDNA; 378 BP.
XX
AC AAK62195;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:7255.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR P-PSDB; AAM89414.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Claim 1; SEQ ID NO 7255; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 378 BP; 135 A; 76 C; 48 G; 116 T; 0 U; 3 Other;

Query Match 100.0%; Score 12; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 227 AAGGTTATATAT 216

Search completed: December 20, 2005, 05:54:36
Job time : 10.2289 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:12:49 ; Search time 2.68273 Seconds
(without alignments)
7951.129 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		12	100.0	12	3	US-08-098-327E-42	Sequence 42, Appl
2		12	100.0	12	3	US-08-098-327E-46	Sequence 46, Appl
3		12	100.0	12	3	US-08-462-625-42	Sequence 42, Appl
4		12	100.0	12	3	US-08-462-625-46	Sequence 46, Appl
c 5		12	100.0	146	3	US-09-513-999C-23604	Sequence 23604, A
c 6		12	100.0	184	3	US-09-480-921B-13	Sequence 13, Appl
c 7		12	100.0	242	3	US-09-513-999C-9073	Sequence 9073, Ap
c 8		12	100.0	294	3	US-09-513-999C-22577	Sequence 22577, A
9		12	100.0	362	3	US-09-513-999C-28463	Sequence 28463, A
c 10		12	100.0	502	3	US-10-002-623-775	Sequence 775, App
c 11		12	100.0	601	3	US-09-949-016-24319	Sequence 24319, A
c 12		12	100.0	601	3	US-09-949-016-24320	Sequence 24320, A
c 13		12	100.0	601	3	US-09-949-016-24321	Sequence 24321, A
c 14		12	100.0	601	3	US-09-949-016-25053	Sequence 25053, A
c 15		12	100.0	601	3	US-09-949-016-25054	Sequence 25054, A
16		12	100.0	601	3	US-09-949-016-28868	Sequence 28868, A
17		12	100.0	601	3	US-09-949-016-28869	Sequence 28869, A
c 18		12	100.0	601	3	US-09-949-016-32135	Sequence 32135, A
c 19		12	100.0	601	3	US-09-949-016-32136	Sequence 32136, A
c 20		12	100.0	601	3	US-09-949-016-32137	Sequence 32137, A
21		12	100.0	601	3	US-09-949-016-32922	Sequence 32922, A
22		12	100.0	601	3	US-09-949-016-32923	Sequence 32923, A
c 23		12	100.0	601	3	US-09-949-016-38188	Sequence 38188, A
c 24		12	100.0	601	3	US-09-949-016-60073	Sequence 60073, A
c 25		12	100.0	601	3	US-09-949-016-68392	Sequence 68392, A
c 26		12	100.0	601	3	US-09-949-016-74492	Sequence 74492, A
27		12	100.0	601	3	US-09-949-016-76566	Sequence 76566, A
28		12	100.0	601	3	US-09-949-016-80457	Sequence 80457, A
c 29		12	100.0	601	3	US-09-949-016-87731	Sequence 87731, A
c 30		12	100.0	601	3	US-09-949-016-94872	Sequence 94872, A
31		12	100.0	601	3	US-09-949-016-115481	Sequence 115481,
32		12	100.0	601	3	US-09-949-016-119463	Sequence 119463,
33		12	100.0	601	3	US-09-949-016-119464	Sequence 119464,
c 34		12	100.0	601	3	US-09-949-016-122654	Sequence 122654,
c 35		12	100.0	601	3	US-09-949-016-125367	Sequence 125367,
c 36		12	100.0	601	3	US-09-949-016-125368	Sequence 125368,
37		12	100.0	601	3	US-09-949-016-130126	Sequence 130126,
38		12	100.0	601	3	US-09-949-016-130127	Sequence 130127,
39		12	100.0	601	3	US-09-949-016-132207	Sequence 132207,
40		12	100.0	601	3	US-09-949-016-132208	Sequence 132208,
41		12	100.0	601	3	US-09-949-016-132275	Sequence 132275,
42		12	100.0	601	3	US-09-949-016-153200	Sequence 153200,
43		12	100.0	601	3	US-09-949-016-153201	Sequence 153201,
c 44		12	100.0	601	3	US-09-949-016-154936	Sequence 154936,
c 45		12	100.0	601	3	US-09-949-016-154937	Sequence 154937,

ALIGNMENTS

```
RESULT 1
US-08-098-327E-42
; Sequence 42, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
```

```
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
US-08-098-327E-46
; Sequence 46, Application US/08098327E
; Patent No. 6270771
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/098,327E
;      FILING DATE:  24-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-045
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-098-327E-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
US-08-462-625-42
; Sequence 42, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
US-08-462-625-46
; Sequence 46, Application US/08462625
; Patent No. 6319502
;  GENERAL INFORMATION:
;    APPLICANT:  GUERIN-MARCHAND, Claudine
;    APPLICANT:  DRUILHE, Pierre
```

```
;    TITLE OF INVENTION:  PEPTIDE SEQUENCES SPECIFIC FOR THE
;    TITLE OF INVENTION:  HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;    TITLE OF INVENTION:  OF STIMULATING THE T LYMPHOCYTES
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,625
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/098,327
;      FILING DATE:  24-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91 01286
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 92/13884
;      PUBLICATION DATE:  20-AUG-1992
US-08-462-625-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
US-09-513-999C-23604/c
; Sequence 23604, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 23604
;   LENGTH: 146
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-23604

  Query Match             100.0%;  Score 12;  DB 3;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         90 AAGGTTATATAT 79


RESULT 6
US-09-480-921B-13/c
; Sequence 13, Application US/09480921B
; Patent No. 6387637
; GENERAL INFORMATION:
;  APPLICANT: Levin, Joshua Z.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Potter, Sharon L.
;  APPLICANT:  Wegrich, Lynette M.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-30780A
;  CURRENT APPLICATION NUMBER: US/09/480,921B
;  CURRENT FILING DATE:  2000-01-11
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 184
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-480-921B-13

  Query Match             100.0%;  Score 12;  DB 3;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 7
US-09-513-999C-9073/c
; Sequence 9073, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 9073
;   LENGTH: 242
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 17
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: 28
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 37
;   OTHER INFORMATION: r=a or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 39
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 45
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 85
;   OTHER INFORMATION: k=g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 106
;   OTHER INFORMATION: b=c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 114
;   OTHER INFORMATION: w=a or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 156
;   OTHER INFORMATION: d=a or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 160
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 187
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 200
;   OTHER INFORMATION: m=a or c
US-09-513-999C-9073

  Query Match             100.0%;  Score 12;  DB 3;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        173 AAGGTTATATAT 162


RESULT 8
US-09-513-999C-22577/c
; Sequence 22577, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 22577
;   LENGTH: 294
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 33
;   OTHER INFORMATION: w=a or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 119
;   OTHER INFORMATION: m=a or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 195
;   OTHER INFORMATION: k=g or t
US-09-513-999C-22577

  Query Match             100.0%;  Score 12;  DB 3;  Length 294;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        240 AAGGTTATATAT 229


RESULT 9
US-09-513-999C-28463
; Sequence 28463, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 28463
;   LENGTH: 362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-28463

  Query Match             100.0%;  Score 12;  DB 3;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        295 AAGGTTATATAT 306


RESULT 10
US-10-002-623-775/c
; Sequence 775, Application US/10002623
; Patent No. 6929911
; GENERAL INFORMATION:
;  APPLICANT: OEFNER, PETER J.
;  APPLICANT:  UNDERHILL, PETER A.
;  TITLE OF INVENTION: A METHOD FOR DETERMINING GENETIC
;  TITLE OF INVENTION:  AFFILIATION, SUBSTRUCTURE AND GENE FLOW WITHIN HUMAN
;  TITLE OF INVENTION:  POPULATIONS
;  FILE REFERENCE: STAN-212
;  CURRENT APPLICATION NUMBER: US/10/002,623
;  CURRENT FILING DATE:  2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/245,355
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 952
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 775
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-002-623-775

  Query Match 100.0%; Score 12; DB 3; Length 502;
  Best Local Similarity 100.0%; Pred. No. 1.2e+03;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
      ||||||||||||
Db 83 AAGGTTATATAT 72

RESULT 11
US-09-949-016-24319/c
; Sequence 24319, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24319
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24319

  Query Match 100.0%; Score 12; DB 3; Length 601;
  Best Local Similarity 100.0%; Pred. No. 1.2e+03;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
      ||||||||||||
Db 595 AAGGTTATATAT 584

RESULT 12
US-09-949-016-24320/c
; Sequence 24320, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24320
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24320

  Query Match 100.0%; Score 12; DB 3; Length 601;
  Best Local Similarity 100.0%; Pred. No. 1.2e+03;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
      ||||||||||||
Db 557 AAGGTTATATAT 546

RESULT 13
US-09-949-016-24321/c
; Sequence 24321, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24321
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24321

  Query Match 100.0%; Score 12; DB 3; Length 601;
  Best Local Similarity 100.0%; Pred. No. 1.2e+03;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
      ||||||||||||
Db 23 AAGGTTATATAT 12

RESULT 14
US-09-949-016-25053/c
; Sequence 25053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25053
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-25053

  Query Match 100.0%; Score 12; DB 3; Length 601;
  Best Local Similarity 100.0%; Pred. No. 1.2e+03;
  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        564 AAGGTTATATAT 553


RESULT 15
US-09-949-016-25054/c
; Sequence 25054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25054
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-25054

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        216 AAGGTTATATAT 205


RESULT 16
US-09-949-016-28868
; Sequence 28868, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28868
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-28868

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        349 AAGGTTATATAT 360


RESULT 17
US-09-949-016-28869
; Sequence 28869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28869
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-28869

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        183 AAGGTTATATAT 194


RESULT 18
US-09-949-016-32135/c
; Sequence 32135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32135
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32135

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        358 AAGGTTATATAT 347


RESULT 19
US-09-949-016-32136/c
; Sequence 32136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32136
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32136

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        140 AAGGTTATATAT 129


RESULT 20
US-09-949-016-32137/c
; Sequence 32137, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32137
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32137

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         62 AAGGTTATATAT 51


RESULT 21
US-09-949-016-32922
; Sequence 32922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32922
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32922

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        244 AAGGTTATATAT 255


RESULT 22
US-09-949-016-32923
; Sequence 32923, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32923
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-32923

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 23
US-09-949-016-38188/c
; Sequence 38188, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
```

```
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38188
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-38188

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        575 AAGGTTATATAT 564


RESULT 24
US-09-949-016-60073/c
; Sequence 60073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60073
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-60073

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        196 AAGGTTATATAT 185


RESULT 25
US-09-949-016-68392/c
; Sequence 68392, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68392
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-68392

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        575 AAGGTTATATAT 564


RESULT 26
US-09-949-016-74492/c
; Sequence 74492, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74492
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-74492

  Query Match             100.0%;  Score 12;  DB 3;  Length 601;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        356 AAGGTTATATAT 345


RESULT 27
US-09-949-016-76566
; Sequence 76566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 76566
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-76566
```

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        339 AAGGTTATATAT 350
```

RESULT 28
US-09-949-016-80457
; Sequence 80457, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 80457
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-80457

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        164 AAGGTTATATAT 175
```

RESULT 29
US-09-949-016-87731/c
; Sequence 87731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 87731
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-87731

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        207 AAGGTTATATAT 196
```

RESULT 30
US-09-949-016-94872/c
; Sequence 94872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 94872
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-94872

Query Match 100.0%; Score 12; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        359 AAGGTTATATAT 348
```

Search completed: December 20, 2005, 15:19:21
Job time : 2.68273 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 10:43:25 ; Search time 13 Seconds
(without alignments)
7633.272 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	12	3	US-09-837-344-42	Sequence 42, Appl
	2	12	100.0	12	3	US-09-837-344-46	Sequence 46, Appl
	3	12	100.0	12	3	US-09-900-963-42	Sequence 42, Appl
	4	12	100.0	12	3	US-09-900-963-46	Sequence 46, Appl
	5	12	100.0	13	8	US-10-257-017B-193201	Sequence 193201,
c	6	12	100.0	13	8	US-10-257-017B-193202	Sequence 193202,
c	7	12	100.0	25	8	US-10-719-900-346370	Sequence 346370,
	8	12	100.0	100	6	US-10-029-386-22059	Sequence 22059, A
	9	12	100.0	146	6	US-10-029-386-14841	Sequence 14841, A
	10	12	100.0	176	3	US-09-969-373-199	Sequence 199, App
	11	12	100.0	176	3	US-09-969-373-200	Sequence 200, App
c	12	12	100.0	184	5	US-10-047-412A-13	Sequence 13, Appl
	13	12	100.0	188	3	US-09-754-853A-298	Sequence 298, App
c	14	12	100.0	190	8	US-10-674-124A-18049	Sequence 18049, A
c	15	12	100.0	201	8	US-10-719-993-30163	Sequence 30163, A
	16	12	100.0	201	8	US-10-741-600-19082	Sequence 19082, A
	17	12	100.0	201	8	US-10-741-600-45291	Sequence 45291, A
c	18	12	100.0	218	7	US-10-424-599-81702	Sequence 81702, A
c	19	12	100.0	232	8	US-10-674-124A-24402	Sequence 24402, A
c	20	12	100.0	242	8	US-10-425-115-54924	Sequence 54924, A
c	21	12	100.0	245	8	US-10-425-115-111652	Sequence 111652,
c	22	12	100.0	250	4	US-09-925-065A-650047	Sequence 650047,
	23	12	100.0	255	3	US-09-867-701-136	Sequence 136, App
	24	12	100.0	267	3	US-09-864-408A-4589	Sequence 4589, Ap
c	25	12	100.0	277	8	US-10-357-930-7109	Sequence 7109, Ap
	26	12	100.0	284	7	US-10-437-963-24548	Sequence 24548, A
c	27	12	100.0	289	3	US-09-867-701-64	Sequence 64, Appl
c	28	12	100.0	313	3	US-09-867-701-3452	Sequence 3452, Ap
c	29	12	100.0	313	8	US-10-674-124A-23482	Sequence 23482, A
	30	12	100.0	316	3	US-09-920-300A-1741	Sequence 1741, Ap
	31	12	100.0	316	5	US-10-033-528-1741	Sequence 1741, Ap
	32	12	100.0	316	6	US-10-099-926-1741	Sequence 1741, Ap
	33	12	100.0	316	9	US-10-961-527-1741	Sequence 1741, Ap
	34	12	100.0	322	8	US-10-425-115-138506	Sequence 138506,
	35	12	100.0	328	8	US-10-674-124A-13190	Sequence 13190, A
c	36	12	100.0	334	4	US-09-925-065A-115836	Sequence 115836,
c	37	12	100.0	334	7	US-10-242-535A-16001	Sequence 16001, A
c	38	12	100.0	334	7	US-10-085-783A-16001	Sequence 16001, A
c	39	12	100.0	337	7	US-10-242-535A-35808	Sequence 35808, A
c	40	12	100.0	337	7	US-10-085-783A-35808	Sequence 35808, A
	41	12	100.0	367	5	US-10-106-698-2914	Sequence 2914, Ap
c	42	12	100.0	373	8	US-10-425-115-27009	Sequence 27009, A
	43	12	100.0	374	3	US-09-803-719-930	Sequence 930, App
	44	12	100.0	374	9	US-10-779-543-14200	Sequence 14200, A
c	45	12	100.0	383	7	US-10-242-535A-17171	Sequence 17171, A

ALIGNMENTS

```
RESULT 1
US-09-837-344-42
; Sequence 42, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
```

```
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-837-344-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 2
US-09-837-344-46
; Sequence 46, Application US/09837344
; Patent No. US20020041882A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/837,344
;             FILING DATE: 19-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/462,625
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: FR 91 01286
;             FILING DATE: 05-FEB-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-078
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 46:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-837-344-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 3
US-09-900-963-42
; Sequence 42, Application US/09900963
; Publication No. US20030064075A1
;   GENERAL INFORMATION:
;        APPLICANT: GUERIN-MARCHAND, Claudine
;                   DRUILHE, Pierre
;        TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                            HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                            OF STIMULATING THE T LYMPHOCYTES
;        NUMBER OF SEQUENCES: 46
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/900,963
;             FILING DATE: 10-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/098,327
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McGowan, Malcolm K.
;             REGISTRATION NUMBER: 39,300
;             REFERENCE/DOCKET NUMBER: 010830-045
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..12
;        PUBLICATION INFORMATION:
;             DOCUMENT NUMBER:  WO 92/13884
;             PUBLICATION DATE:  20-AUG-1992
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-900-963-42

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 4
US-09-900-963-46
```

```
; Sequence 46, Application US/09900963
; Publication No. US20030064075A1
;  GENERAL INFORMATION:
;       APPLICANT: GUERIN-MARCHAND, Claudine
;                  DRUILHE, Pierre
;       TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
;                           HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
;                           OF STIMULATING THE T LYMPHOCYTES
;       NUMBER OF SEQUENCES: 46
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/900,963
;            FILING DATE: 10-Jul-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/098,327
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McGowan, Malcolm K.
;            REGISTRATION NUMBER: 39,300
;            REFERENCE/DOCKET NUMBER: 010830-045
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 12 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..12
;       PUBLICATION INFORMATION:
;            DOCUMENT NUMBER:  WO 92/13884
;            PUBLICATION DATE:  20-AUG-1992
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-900-963-46

  Query Match             100.0%;  Score 12;  DB 3;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 5
US-10-257-017B-193201
; Sequence 193201, Application US/10257017B
; Publication No. US20040241651A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Detection of single nucleotide polymorhphisms [SNPs] and cytosine
;  TITLE OF INVENTION:  methylations
;  FILE REFERENCE: E01/1193/WO
;  CURRENT APPLICATION NUMBER: US/10/257,017B
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 382046
; SEQ ID NO 193201
;   LENGTH: 13
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide for detection of SNP TSC0047525
US-10-257-017B-193201

  Query Match             100.0%;  Score 12;  DB 8;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          1 AAGGTTATATAT 12


RESULT 6
US-10-257-017B-193202/c
; Sequence 193202, Application US/10257017B
; Publication No. US20040241651A1
; GENERAL INFORMATION:
;  APPLICANT: Alexander Olek
;  APPLICANT:  Christian Piepenbrock
;  APPLICANT:  Kurt Berlin
;  TITLE OF INVENTION: Detection of single nucleotide polymorhphisms [SNPs] and cytosine
;  TITLE OF INVENTION:  methylations
;  FILE REFERENCE: E01/1193/WO
;  CURRENT APPLICATION NUMBER: US/10/257,017B
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 382046
; SEQ ID NO 193202
;   LENGTH: 13
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide for detection of SNP TSC0047525
US-10-257-017B-193202

  Query Match             100.0%;  Score 12;  DB 8;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 9.1e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         13 AAGGTTATATAT 2


RESULT 7
US-10-719-900-346370/c
; Sequence 346370, Application US/10719900
; Publication No. US20050026164A1
; GENERAL INFORMATION:
;  APPLICANT: Xue Mei Zhou
;  TITLE OF INVENTION: Methods of Genetic Analysis of Mouse
;  FILE REFERENCE: 3528.1
;  CURRENT APPLICATION NUMBER: US/10/719,900
;  CURRENT FILING DATE:  2003-11-20
;  PRIOR APPLICATION NUMBER: 60/427,808
;  PRIOR FILING DATE: 2002 11 20
;  NUMBER OF SEQ ID NOS: 982914
;  SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 346370
;   LENGTH: 25
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-719-900-346370
```

```
  Query Match           100.0%;  Score 12;  DB 8;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         20 AAGGTTATATAT 9


RESULT 8
US-10-029-386-22059
; Sequence 22059, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22059
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR14.3
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
;   OTHER INFORMATION: NT HIT: AB000381.1, EVALUE 8.00e-02
;   OTHER INFORMATION: EST_HUMAN HIT: H08306.1, EVALUE 4.70e-01
;   OTHER INFORMATION: SWISSPROT HIT: P30796, EVALUE 6.10e+00
US-10-029-386-22059

  Query Match           100.0%;  Score 12;  DB 6;  Length 100;
  Best Local Similarity  100.0%;  Pred. No. 9.9e+03;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         44 AAGGTTATATAT 55


RESULT 9
US-10-029-386-14841
; Sequence 14841, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14841
;   LENGTH: 146
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL121957.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.57
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.45
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.44
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.42
;   OTHER INFORMATION: EST_HUMAN HIT: BG958104.1, EVALUE 7.30e-01
;   OTHER INFORMATION: NT HIT: gi15295339, EVALUE 5.00e-01
US-10-029-386-14841

  Query Match           100.0%;  Score 12;  DB 6;  Length 146;
  Best Local Similarity  100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         97 AAGGTTATATAT 108


RESULT 10
US-09-969-373-199
; Sequence 199, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 199
;   LENGTH: 176
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-199

  Query Match           100.0%;  Score 12;  DB 3;  Length 176;
  Best Local Similarity  100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         21 AAGGTTATATAT 32


RESULT 11
US-09-969-373-200
; Sequence 200, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 200
;   LENGTH: 176
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-200

  Query Match           100.0%;  Score 12;  DB 3;  Length 176;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         21 AAGGTTATATAT 32


RESULT 12
US-10-047-412A-13/c
; Sequence 13, Application US/10047412A
; Publication No. US20020197696A1
; GENERAL INFORMATION:
;  APPLICANT: Levin, Joshua Z.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Potter, Sharon L.
;  APPLICANT:  Wegrich, Lynette M.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-30780DIV
;  CURRENT APPLICATION NUMBER: US/10/047,412A
;  CURRENT FILING DATE:  2002-04-11
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 184
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-10-047-412A-13

  Query Match             100.0%;  Score 12;  DB 5;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         86 AAGGTTATATAT 75


RESULT 13
US-09-754-853A-298
; Sequence 298, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 298
;   LENGTH: 188
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 318O13_region_A3__153114_12
US-09-754-853A-298

  Query Match             100.0%;  Score 12;  DB 3;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         31 AAGGTTATATAT 42


RESULT 14
US-10-674-124A-18049/c
; Sequence 18049, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 18049
;   LENGTH: 190
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: D11S1340
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 11
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 123440912
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 149287
US-10-674-124A-18049

  Query Match             100.0%;  Score 12;  DB 8;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        164 AAGGTTATATAT 153


RESULT 15
US-10-719-993-30163/c
; Sequence 30163, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30163
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-30163

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
```

```
              ||||||||||||
Db         63 AAGGTTATATAT 52


RESULT 16
US-10-741-600-19082
; Sequence 19082, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19082
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-19082

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         38 AAGGTTATATAT 49


RESULT 17
US-10-741-600-45291
; Sequence 45291, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45291
;   LENGTH: 201
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-45291

  Query Match             100.0%;  Score 12;  DB 8;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         73 AAGGTTATATAT 84


RESULT 18
US-10-424-599-81702/c
; Sequence 81702, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 81702
;   LENGTH: 218
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(218)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_44794C.1
US-10-424-599-81702

  Query Match             100.0%;  Score 12;  DB 7;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         28 AAGGTTATATAT 17


RESULT 19
US-10-674-124A-24402/c
; Sequence 24402, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 24402
;   LENGTH: 232
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: chr20.fa.07frz.6706738
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 20
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 6258865
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 90016
US-10-674-124A-24402

  Query Match             100.0%;  Score 12;  DB 8;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         96 AAGGTTATATAT 85
```

```
RESULT 20
US-10-425-115-54924/c
; Sequence 54924, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 54924
;   LENGTH: 242
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(242)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_150089C.1
US-10-425-115-54924

  Query Match             100.0%;  Score 12;  DB 8;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        187 AAGGTTATATAT 176


RESULT 21
US-10-425-115-111652/c
; Sequence 111652, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 111652
;   LENGTH: 245
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_33313C.1
US-10-425-115-111652

  Query Match             100.0%;  Score 12;  DB 8;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        226 AAGGTTATATAT 215


RESULT 22
US-09-925-065A-650047/c
; Sequence 650047, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 650047
;   LENGTH: 250
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-650047

  Query Match             100.0%;  Score 12;  DB 4;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         51 AAGGTTATATAT 40


RESULT 23
US-09-867-701-136
; Sequence 136, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 136
;   LENGTH: 255
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-136

  Query Match             100.0%;  Score 12;  DB 3;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         26 AAGGTTATATAT 37


RESULT 24
US-09-864-408A-4589
; Sequence 4589, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Shimkets, Richard A.
;  TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Encc
;  FILE REFERENCE: 21402-012
;  CURRENT APPLICATION NUMBER: US/09/864,408A
```

```
;  CURRENT FILING DATE:  2001-05-24
;  PRIOR APPLICATION NUMBER: 60/206,690
;  PRIOR FILING DATE: 2000-05-24
;  NUMBER OF SEQ ID NOS: 9068
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4589
;   LENGTH: 267
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (13)..(13)
;   OTHER INFORMATION: Wherein n may be a, c, g or t
US-09-864-408A-4589

  Query Match            100.0%;  Score 12;  DB 3;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         23 AAGGTTATATAT 34


RESULT 25
US-10-357-930-7109/c
; Sequence 7109, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007BCN
;  CURRENT APPLICATION NUMBER: US/10/357,930
;  CURRENT FILING DATE:  2003-02-04
;  PRIOR APPLICATION NUMBER: 09/785,276
;  PRIOR FILING DATE: 2003-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7109
;   LENGTH: 277
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 90, 257
;   OTHER INFORMATION: n = A,T,C or G
US-10-357-930-7109

  Query Match            100.0%;  Score 12;  DB 8;  Length 277;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         76 AAGGTTATATAT 65


RESULT 26
US-10-437-963-24548
; Sequence 24548, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24548
;   LENGTH: 284
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_29521C.1
US-10-437-963-24548

  Query Match            100.0%;  Score 12;  DB 7;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        178 AAGGTTATATAT 189


RESULT 27
US-09-867-701-64/c
; Sequence 64, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Aglate, Paul A.
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.497
;  CURRENT APPLICATION NUMBER: US/09/867,701
;  CURRENT FILING DATE:  2001-05-29
;  NUMBER OF SEQ ID NOS: 10912
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 64
;   LENGTH: 289
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-867-701-64

  Query Match            100.0%;  Score 12;  DB 3;  Length 289;
  Best Local Similarity   100.0%;  Pred. No. 1e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        250 AAGGTTATATAT 239


RESULT 28
US-09-867-701-3452/c
; Sequence 3452, Application US/09867701
```

; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3452
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(313)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-3452

Query Match 100.0%; Score 12; DB 3; Length 313;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 303 AAGGTTATATAT 292

RESULT 29
US-10-674-124A-23482/c
; Sequence 23482, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hidetoshi
; APPLICANT: TAMIYA, Gen
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; TITLE OF INVENTION: GENETIC POLYMORPHISM MARKERS
; FILE REFERENCE: ORIN-003CIP
; CURRENT APPLICATION NUMBER: US/10/674,124A
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 23482
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: AC067801.2_170716
; FEATURE:
; OTHER INFORMATION: Located on chromosome 18
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; OTHER INFORMATION: sequence : 43376825
; FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 137159
US-10-674-124A-23482

Query Match 100.0%; Score 12; DB 8; Length 313;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 224 AAGGTTATATAT 213

RESULT 30
US-09-920-300A-1741
; Sequence 1741, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1741
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13, 22, 237, 298
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-1741

Query Match 100.0%; Score 12; DB 3; Length 316;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 161 AAGGTTATATAT 172

Search completed: December 20, 2005, 23:02:14
Job time : 13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 15:08:19 ; Search time 3.30924 Seconds
(without alignments)
1777.494 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161431 seqs, 245089505 residues

Total number of hits satisfying chosen parameters: 8322862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	12	100.0	201	6	US-10-995-561-14708	Sequence 14708, A
c	2	12	100.0	600	6	US-10-750-185-1498	Sequence 1498, Ap
c	3	12	100.0	605	6	US-10-750-185-658	Sequence 658, App
c	4	12	100.0	802	6	US-10-750-185-33511	Sequence 33511, A
	5	12	100.0	963	6	US-10-750-185-28663	Sequence 28663, A
c	6	12	100.0	1026	6	US-10-750-185-33249	Sequence 33249, A
c	7	12	100.0	1174	6	US-10-750-185-52955	Sequence 52955, A
	8	12	100.0	1327	6	US-10-750-185-41599	Sequence 41599, A
	9	12	100.0	1332	6	US-10-750-185-62132	Sequence 62132, A
c	10	12	100.0	1404	6	US-10-750-185-26764	Sequence 26764, A
	11	12	100.0	1439	6	US-10-750-185-52757	Sequence 52757, A
	12	12	100.0	1650	6	US-10-750-185-54339	Sequence 54339, A
c	13	12	100.0	1913	6	US-10-750-185-27237	Sequence 27237, A
	14	12	100.0	2004	6	US-10-750-185-58707	Sequence 58707, A
	15	12	100.0	2050	6	US-10-750-185-37205	Sequence 37205, A
c	16	12	100.0	2269	7	US-11-087-072-1	Sequence 1, Appli
	17	12	100.0	2474	6	US-10-750-185-31559	Sequence 31559, A
c	18	12	100.0	2685	6	US-10-750-185-39093	Sequence 39093, A
	19	12	100.0	2688	6	US-10-750-185-55769	Sequence 55769, A
	20	12	100.0	3118	6	US-10-793-626-3939	Sequence 3939, Ap
	21	12	100.0	3187	6	US-10-793-626-4381	Sequence 4381, Ap
c	22	12	100.0	3320	6	US-10-909-125-797	Sequence 797, App
	23	12	100.0	3712	6	US-10-750-185-61739	Sequence 61739, A
c	24	12	100.0	4880	6	US-10-750-185-31043	Sequence 31043, A
c	25	12	100.0	6192	6	US-10-750-185-44297	Sequence 44297, A
	26	12	100.0	13363	6	US-10-995-561-13446	Sequence 13446, A
	27	12	100.0	29959	6	US-10-995-561-13475	Sequence 13475, A
c	28	12	100.0	33019	6	US-10-995-561-13371	Sequence 13371, A
c	29	12	100.0	57917	6	US-10-995-561-13299	Sequence 13299, A
	30	12	100.0	95832	6	US-10-995-561-13273	Sequence 13273, A
c	31	12	100.0	103931	7	US-11-117-187-193	Sequence 193, App
c	32	12	100.0	149419	7	US-11-112-908-49	Sequence 49, Appl
c	33	12	100.0	161726	7	US-11-112-908-48	Sequence 48, Appl
c	34	12	100.0	161726	7	US-11-112-908-52	Sequence 52, Appl
c	35	12	100.0	166111	7	US-11-112-908-47	Sequence 47, Appl
	36	12	100.0	180654	7	US-11-121-086-58	Sequence 58, Appl
c	37	12	100.0	185393	7	US-11-121-086-101	Sequence 101, App
c	38	12	100.0	187745	7	US-11-121-086-83	Sequence 83, Appl
c	39	12	100.0	188682	7	US-11-112-908-23	Sequence 23, Appl
c	40	12	100.0	191797	7	US-11-121-086-13	Sequence 13, Appl
	41	12	100.0	209822	6	US-10-995-561-13198	Sequence 13198, A
	42	12	100.0	212716	7	US-11-121-086-95	Sequence 95, Appl
c	43	12	100.0	212716	7	US-11-121-086-95	Sequence 95, Appl
c	44	12	100.0	319608	7	US-11-145-703-1	Sequence 1, Appli
	45	12	100.0	387780	6	US-10-995-561-13259	Sequence 13259, A

ALIGNMENTS

RESULT 1
US-10-995-561-14708
; Sequence 14708, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14708
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-14708

Query Match 100.0%; Score 12; DB 6; Length 201;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         38 AAGGTTATATAT 49
```

RESULT 2
US-10-750-185-1498/c
; Sequence 1498, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482

```
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 1498
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bovine  MMBT17112
US-10-750-185-1498

  Query Match            100.0%;  Score 12;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        553 AAGGTTATATAT 542


RESULT 3
US-10-750-185-658/c
; Sequence 658, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 658
;   LENGTH: 605
;   TYPE: DNA
;   ORGANISM: Bovine  MMBT03522
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(5)
;   OTHER INFORMATION: n is any nucleotide
US-10-750-185-658

  Query Match            100.0%;  Score 12;  DB 6;  Length 605;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        345 AAGGTTATATAT 334


RESULT 4
US-10-750-185-33511/c
; Sequence 33511, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33511
;   LENGTH: 802
;   TYPE: DNA
;   ORGANISM: Bovine  19866880419621
US-10-750-185-33511

  Query Match            100.0%;  Score 12;  DB 6;  Length 802;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        631 AAGGTTATATAT 620


RESULT 5
US-10-750-185-28663
; Sequence 28663, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 28663
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Bovine  19866880781472
US-10-750-185-28663

  Query Match            100.0%;  Score 12;  DB 6;  Length 963;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         54 AAGGTTATATAT 65


RESULT 6
US-10-750-185-33249/c
; Sequence 33249, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
```

```
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33249
;   LENGTH: 1026
;   TYPE: DNA
;   ORGANISM: Bovine  19866881769644
US-10-750-185-33249

  Query Match             100.0%;  Score 12;  DB 6;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        162 AAGGTTATATAT 151


RESULT 7
US-10-750-185-52955/c
; Sequence 52955, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52955
;   LENGTH: 1174
;   TYPE: DNA
;   ORGANISM: Bovine  19866881042220
US-10-750-185-52955

  Query Match             100.0%;  Score 12;  DB 6;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         29 AAGGTTATATAT 18


RESULT 8
US-10-750-185-41599
; Sequence 41599, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41599
;   LENGTH: 1327
;   TYPE: DNA
;   ORGANISM: Bovine  19866881435791
US-10-750-185-41599

  Query Match             100.0%;  Score 12;  DB 6;  Length 1327;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        944 AAGGTTATATAT 955


RESULT 9
US-10-750-185-62132
; Sequence 62132, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 62132
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Bovine  19866880526631
US-10-750-185-62132

  Query Match             100.0%;  Score 12;  DB 6;  Length 1332;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        411 AAGGTTATATAT 422


RESULT 10
US-10-750-185-26764/c
; Sequence 26764, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 26764
;   LENGTH: 1404
;   TYPE: DNA
;   ORGANISM: Bovine  19866881239706
US-10-750-185-26764
```

```
  Query Match             100.0%;  Score 12;  DB 6;  Length 1404;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        965 AAGGTTATATAT 954


RESULT 11
US-10-750-185-52757
; Sequence 52757, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52757
;   LENGTH: 1439
;   TYPE: DNA
;   ORGANISM: Bovine  19866880951582
US-10-750-185-52757

  Query Match             100.0%;  Score 12;  DB 6;  Length 1439;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        455 AAGGTTATATAT 466


RESULT 12
US-10-750-185-54339
; Sequence 54339, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 54339
;   LENGTH: 1650
;   TYPE: DNA
;   ORGANISM: Bovine  19866880458349
US-10-750-185-54339

  Query Match             100.0%;  Score 12;  DB 6;  Length 1650;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        341 AAGGTTATATAT 352


RESULT 13
US-10-750-185-27237/c
; Sequence 27237, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 27237
;   LENGTH: 1913
;   TYPE: DNA
;   ORGANISM: Bovine  19866880640589
US-10-750-185-27237

  Query Match             100.0%;  Score 12;  DB 6;  Length 1913;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        338 AAGGTTATATAT 327


RESULT 14
US-10-750-185-58707
; Sequence 58707, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 58707
;   LENGTH: 2004
;   TYPE: DNA
;   ORGANISM: Bovine  19866880608793
US-10-750-185-58707

  Query Match             100.0%;  Score 12;  DB 6;  Length 2004;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
```

```
              ||||||||||||
Db        793 AAGGTTATATAT 804


RESULT 15
US-10-750-185-37205
; Sequence 37205, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 37205
;   LENGTH: 2050
;   TYPE: DNA
;   ORGANISM: Bovine  19866880646640
US-10-750-185-37205

  Query Match             100.0%;  Score 12;  DB 6;  Length 2050;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1771 AAGGTTATATAT 1782


RESULT 16
US-11-087-072-1/c
; Sequence 1, Application US/11087072
; Publication No. US20050272069A1
;   GENERAL INFORMATION:
;        APPLICANT: Olopade, Olufunmilayo I.
;        TITLE OF INVENTION: METHYLTHIOADENOSINE PHOSPHORYLASE
;                            COMPOSITIONS AND METHODS OF USE IN
;                            THE DIAGNOSIS AND TREATMENT OF
;                            PROLIFERATIVE DISORDERS
;        NUMBER OF SEQUENCES: 15
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: United States of America
;             ZIP: 77210
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/11/087,072
;             FILING DATE: 22-Mar-2005
;             CLASSIFICATION: 435
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/674,311
;             FILING DATE: 01-JUL-1996
;             APPLICATION NUMBER: US 60/000,831
;             FILING DATE: 02-JUL-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kitchell, Barbara S.
;             REGISTRATION NUMBER: 33,928
;             REFERENCE/DOCKET NUMBER: ARSB:509
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2269 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  122..970
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-11-087-072-1

  Query Match             100.0%;  Score 12;  DB 7;  Length 2269;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1776 AAGGTTATATAT 1765


RESULT 17
US-10-750-185-31559
; Sequence 31559, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31559
;   LENGTH: 2474
;   TYPE: DNA
;   ORGANISM: Bovine  19866880685931
US-10-750-185-31559

  Query Match             100.0%;  Score 12;  DB 6;  Length 2474;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1439 AAGGTTATATAT 1450


RESULT 18
US-10-750-185-39093/c
; Sequence 39093, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
```

```
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39093
;   LENGTH: 2685
;   TYPE: DNA
;   ORGANISM: Bovine  19866880416817
US-10-750-185-39093

  Query Match             100.0%;  Score 12;  DB 6;  Length 2685;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        223 AAGGTTATATAT 212


RESULT 19
US-10-750-185-55769
; Sequence 55769, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55769
;   LENGTH: 2688
;   TYPE: DNA
;   ORGANISM: Bovine  19866880540328
US-10-750-185-55769

  Query Match             100.0%;  Score 12;  DB 6;  Length 2688;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       2350 AAGGTTATATAT 2361


RESULT 20
US-10-793-626-3939
; Sequence 3939, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
;   LENGTH: 3118
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3939

  Query Match             100.0%;  Score 12;  DB 6;  Length 3118;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db       1803 AAGGTTATATAT 1814


RESULT 21
US-10-793-626-4381
; Sequence 4381, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4381
;   LENGTH: 3187
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4381

  Query Match             100.0%;  Score 12;  DB 6;  Length 3187;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        593 AAGGTTATATAT 604


RESULT 22
US-10-909-125-797/c
; Sequence 797, Application US/10909125
; Publication No. US20050261218A1
; GENERAL INFORMATION:
;  APPLICANT: Esau, Christine
;  APPLICANT:  Lollo, Bridget
;  APPLICANT:  Bennett, C. Frank
;  APPLICANT:  Freier, Susan M.
;  APPLICANT:  Griffey, Richard H.
;  APPLICANT:  Baker, Brenda F.
;  APPLICANT:  Vickers, Timothy
;  APPLICANT:  Marcusson, Eric G.
;  APPLICANT:  Koller, Erich
;  APPLICANT:  Swayze, Eric
;  APPLICANT:  Jain, Ravi
;  APPLICANT:  Bhat, Balkrishen
;  APPLICANT:  Peralta, Eigen
;  TITLE OF INVENTION: Oligomeric Compounds And Compositions For Use In Modulation
;  TITLE OF INVENTION:  Of Small Non-Coding RNAs
;  FILE REFERENCE: ISIS0080-100 (CORE0016US)
;  CURRENT APPLICATION NUMBER: US/10/909,125
;  CURRENT FILING DATE:  2004-07-30
;  PRIOR APPLICATION NUMBER: US 60/492,056
```

; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: US 60/516,303
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: US 60/531,596
; PRIOR FILING DATE: 2003-12-19
; PRIOR APPLICATION NUMBER: US 60/562,417
; PRIOR FILING DATE: 2004-04-14
; NUMBER OF SEQ ID NOS: 2184
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 797
; LENGTH: 3320
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-909-125-797

Query Match 100.0%; Score 12; DB 6; Length 3320;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 2662 AAGGTTATATAT 2651

RESULT 23
US-10-750-185-61739
; Sequence 61739, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 61739
; LENGTH: 3712
; TYPE: DNA
; ORGANISM: Bovine 19866881126830
US-10-750-185-61739

Query Match 100.0%; Score 12; DB 6; Length 3712;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 2143 AAGGTTATATAT 2154

RESULT 24
US-10-750-185-31043/c
; Sequence 31043, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 31043
; LENGTH: 4880
; TYPE: DNA
; ORGANISM: Bovine 19866880412972
US-10-750-185-31043

Query Match 100.0%; Score 12; DB 6; Length 4880;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 4367 AAGGTTATATAT 4356

RESULT 25
US-10-750-185-44297/c
; Sequence 44297, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MMI1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 44297
; LENGTH: 6192
; TYPE: DNA
; ORGANISM: Bovine 19866881113731
US-10-750-185-44297

Query Match 100.0%; Score 12; DB 6; Length 6192;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 665 AAGGTTATATAT 654

RESULT 26
US-10-995-561-13446
; Sequence 13446, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13446
; LENGTH: 13363
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-10-995-561-13446

  Query Match             100.0%;  Score 12;  DB 6;  Length 13363;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      13052 AAGGTTATATAT 13063


RESULT 27
US-10-995-561-13475
; Sequence 13475, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13475
;   LENGTH: 29959
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13475

  Query Match             100.0%;  Score 12;  DB 6;  Length 29959;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      19791 AAGGTTATATAT 19802


RESULT 28
US-10-995-561-13371/c
; Sequence 13371, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13371
;   LENGTH: 33019
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13371

  Query Match             100.0%;  Score 12;  DB 6;  Length 33019;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      13394 AAGGTTATATAT 13383


RESULT 29
US-10-995-561-13299/c
; Sequence 13299, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13299
;   LENGTH: 57917
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13299

  Query Match             100.0%;  Score 12;  DB 6;  Length 57917;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      55957 AAGGTTATATAT 55946


RESULT 30
US-10-995-561-13273
; Sequence 13273, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13273
;   LENGTH: 95832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13273

  Query Match             100.0%;  Score 12;  DB 6;  Length 95832;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db      76948 AAGGTTATATAT 76959


Search completed: December 20, 2005, 23:16:08
Job time : 6.30924 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:31:02 ; Search time 0.447791 Seconds
(without alignments)
5156.873 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 565948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093216_22943/app_query.fasta_1.3740
-DB=PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_179_@runat_21122005_093216_22943 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 11:33:17
Job time : 0.447791 secs

-

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2005, 05:09:34 ; Search time 63.7751 Seconds
(without alignments)
8803.514 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 12
Sequence: 1 AAGGTTATATAT 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		12	100.0	66	9	CC884412	CC884412 SALK_1129
2		12	100.0	73	10	AL757665	AL757665 Arabidops
3		12	100.0	94	10	BX897322	BX897322 Arabidops
	c 4	12	100.0	99	3	BJ031417	BJ031417 BJ031417
	c 5	12	100.0	102	1	AW066376	AW066376 687001E09
6		12	100.0	102	5	BW423536	BW423536 BW423536
	c 7	12	100.0	106	1	AI943870	AI943870 614030E12
	c 8	12	100.0	109	3	BI847701	BI847701 469886 MA
	c 9	12	100.0	112	1	AI810434	AI810434 wb84f01.x
	c 10	12	100.0	116	9	CC428121	CC428121 PUHBC05TB
	c 11	12	100.0	116	10	BX535978	BX535978 Arabidops
	c 12	12	100.0	117	10	BX535977	BX535977 Arabidops
	c 13	12	100.0	135	10	CL883998	CL883998 abf64a08.
14		12	100.0	137	7	CK923106	CK923106 p3fmgc_01
15		12	100.0	146	5	BX327794	BX327794 BX327794
	c 16	12	100.0	147	9	B95930	B95930 T18I6TF TAM
	c 17	12	100.0	148	6	CA364088	CA364088 638929 NC
18		12	100.0	150	2	BG929548	BG929548 950006A12
19		12	100.0	153	1	AV107977	AV107977 AV107977
	c 20	12	100.0	155	2	BE058406	BE058406 sn15f09.y
21		12	100.0	157	6	CA766087	CA766087 AF53-Rpf_
	c 22	12	100.0	158	7	CO720276	CO720276 tai41f12.
	c 23	12	100.0	159	8	T26907	T26907 ESTDIR575 C
24		12	100.0	160	6	CD162149	CD162149 ML1-0075T
	c 25	12	100.0	161	2	BG513953	BG513953 dac15h11.
26		12	100.0	163	10	CE584959	CE584959 tigr-gss-
	c 27	12	100.0	165	10	BX655471	BX655471 Arabidops
	c 28	12	100.0	166	7	CV314160	CV314160 CM1-FN010
29		12	100.0	168	2	BF911105	BF911105 PM1-UT006
	c 30	12	100.0	168	3	BP616314	BP616314 BP616314
31		12	100.0	168	10	CE297732	CE297732 tigr-gss-
32		12	100.0	169	9	BH559793	BH559793 BOGVD49TR
33		12	100.0	170	1	AV106575	AV106575 AV106575
34		12	100.0	172	9	BH532142	BH532142 BOGVD38TR
35		12	100.0	172	10	CE853534	CE853534 tigr-gss-
36		12	100.0	173	2	BG937368	BG937368 950006A12
37		12	100.0	174	1	AV106563	AV106563 AV106563
38		12	100.0	177	1	BB000380	BB000380 BB000380
	c 39	12	100.0	177	3	BP618736	BP618736 BP618736
40		12	100.0	178	1	AV104185	AV104185 AV104185
	c 41	12	100.0	178	6	CD024554	CD024554 NXRV_044_
42		12	100.0	178	7	CV385592	CV385592 QV1-BT082
43		12	100.0	180	10	CW801296	CW801296 WiscDsLox
44		12	100.0	186	1	AI842967	AI842967 UI-M-AO1-
	c 45	12	100.0	187	9	AZ830376	AZ830376 2M0109P06

ALIGNMENTS

RESULT 1
CC884412
LOCUS CC884412 66 bp DNA linear GSS 31-JUL-2003
DEFINITION SALK_112931.20.95.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_112931.20.95.x, genomic survey sequence.
ACCESSION CC884412
VERSION CC884412.1 GI:33360768
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 66)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At3g13660.
Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .66
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_112931.20.95.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 27 AAGGTTATATAT 38

RESULT 2
AL757665
LOCUS AL757665 73 bp DNA linear GSS 01-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-151D09-013133, genomic survey sequence.
ACCESSION AL757665
VERSION AL757665.1 GI:21496013
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 73)
AUTHORS Rosso,M.G., Li,Y., Strizhov,N. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At2g29840. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .73
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-151D09-013133"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion.

T-DNA derived sequences were removed."
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 73;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 46 AAGGTTATATAT 57

RESULT 3
BX897322
LOCUS BX897322 94 bp DNA linear GSS 05-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-766A12-023593, genomic survey sequence.
ACCESSION BX897322
VERSION BX897322.1 GI:39929817
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 94)
AUTHORS Strizhov,N., Rosso,M.G., Li,Y. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At1g53030. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .94
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-766A12-023593"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion.

T-DNA derived sequences were removed."
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 69 AAGGTTATATAT 80

RESULT 4
BJ031417/c
LOCUS BJ031417 99 bp mRNA linear EST 26-SEP-2003
DEFINITION BJ031417 NIBB Mochii normalized Xenopus neurula library Xenopus laevis cDNA clone XL007b24 5', mRNA sequence.
ACCESSION BJ031417
VERSION BJ031417.1 GI:17408235
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 99)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.
http://xenopus.nibb.ac.jp.
FEATURES Location/Qualifiers
source 1. .99
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL007b24"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/clone_lib="NIBB Mochii normalized Xenopus neurula library"
ORIGIN

Query Match 100.0%; Score 12; DB 3; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 27 AAGGTTATATAT 16

RESULT 5
AW066376/c
LOCUS AW066376 102 bp mRNA linear EST 30-MAR-2000
DEFINITION 687001E09.y1 687 - Early embryo from Delaware Zea mays cDNA, mRNA sequence.
ACCESSION AW066376
VERSION AW066376.1 GI:6021448
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 102)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687001 row: E column: 09.
FEATURES Location/Qualifiers
source 1. .102
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/tissue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/clone_lib="687 - Early embryo from Delaware"
/note="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI; Site_2: EcoRI; Library was prepared by Statagene using the Uni-ZAP XR system (Stratagene BN937328-12). Clones were picked by a Q-bot after blue/white selection (ampicillin resistance - use 100 micrograms/microliter). Developed from a pool of equal amounts of RNA from developing embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize Strain Cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"
ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 97 AAGGTTATATAT 86

RESULT 6
BW423536
LOCUS BW423536 102 bp mRNA linear EST 02-JUN-2004
DEFINITION BW423536 Yutaka Satou unpublished cDNA library, mature adult whole animal Ciona intestinalis cDNA clone cima846f02 3', mRNA sequence.
ACCESSION BW423536
VERSION BW423536.1 GI:47961086
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 102)
AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Yutaka Satou
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES Location/Qualifiers
source 1. .102
/organism="Ciona intestinalis"
/mol_type="mRNA"

```
                     /db_xref="taxon:7719"
                     /clone="cima846f02"
                     /tissue_type="whole animal"
                     /dev_stage="mature adult"
                     /clone_lib="Yutaka Satou unpublished cDNA library, mature
                     adult whole animal"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 5;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         15 AAGGTTATATAT 26


RESULT 7
AI943870/c
LOCUS       AI943870                 106 bp    mRNA    linear   EST 09-AUG-1999
DEFINITION  614030E12.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI943870
VERSION     AI943870.1  GI:5713885
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 106)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614030  row: E  column: 12.
FEATURES             Location/Qualifiers
     source          1. .106
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 1;  Length 106;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        104 AAGGTTATATAT 93


RESULT 8
BI847701/c
LOCUS       BI847701                 109 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  469886 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI847701
VERSION     BI847701.1  GI:15960220
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: L  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .109
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 3;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         70 AAGGTTATATAT 59


RESULT 9
AI810434/c
LOCUS       AI810434                 112 bp    mRNA    linear   EST 17-DEC-1999
DEFINITION  wb84f01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2312377 3',
            mRNA sequence.
ACCESSION   AI810434
VERSION     AI810434.1  GI:5397000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 112)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 469 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 79.

FEATURES Location/Qualifiers
source 1. .112
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2312377"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 100.0%; Score 12; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 41 AAGGTTATATAT 30

RESULT 10
CC428121/c
LOCUS CC428121 116 bp DNA linear GSS 19-MAY-2003
DEFINITION PUHBC05TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa411A09, genomic survey sequence.
ACCESSION CC428121
VERSION CC428121.1 GI:30908211
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 116)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES Location/Qualifiers
source 1. .116
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa411A09"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high CoT selected genomic DNA library"

ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 115 AAGGTTATATAT 104

RESULT 11
BX535978/c
LOCUS BX535978 116 bp DNA linear GSS 04-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-522G02-020204, genomic survey sequence.
ACCESSION BX535978
VERSION BX535978.1 GI:31413108
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 116)
AUTHORS Strizhov,N., Rosso,M.G., Li,Y. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At5g54040. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES Location/Qualifiers
source 1. .116
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-522G02-020204"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from

vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        111 AAGGTTATATAT 100
```

RESULT 12
BX535977/c
LOCUS BX535977 117 bp DNA linear GSS 04-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-522G02-020049, genomic survey sequence.
ACCESSION BX535977
VERSION BX535977.1 GI:31413107
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 117)
AUTHORS Li,Y., Rosso,M.G., Strizhov,N. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At5g54040. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .117
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-522G02-020049"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from

vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        112 AAGGTTATATAT 101
```

RESULT 13
CL883998/c
LOCUS CL883998 135 bp DNA linear GSS 30-AUG-2004
DEFINITION abf64a08.y1 Soybean random, unfiltered genomic library Glycine max genomic, genomic survey sequence.
ACCESSION CL883998
VERSION CL883998.1 GI:51622431
KEYWORDS GSS.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 135)
AUTHORS Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J., Peterson,S., Jones,J., Fries,J., Budiman,M.A., Nguyen,H. and Stacey,G.
TITLE Methylation filtered genomic sequences from Glycine max
JOURNAL Unpublished (2004)
COMMENT Contact: Gary Stacey
University of Missouri
108 Waters Hall, Columbia, MO 65211, USA
Tel: 573-884-1267
Fax: 573-882-0588
Email: staceyg@missouri.edu
LidID: 230
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .135
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Williams 82"
/db_xref="taxon:3847"
/tissue_type="Young leaves"
/clone_lib="Soybean random, unfiltered genomic library"
/note="Vector: pOT2; Site_1: BstXI; Randomly sheared genomic DNA ranging from 0.7-1.5 kb were end repaired and ligated to BstXI linkers prior to cloning in BstXI-cut pOT2. LibID: 230"
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         18 AAGGTTATATAT 7
```

RESULT 14
CK923106
LOCUS CK923106 137 bp mRNA linear EST 22-APR-2004
DEFINITION p3fmgc_012733 Normalized Magnaporthe grisea cDNA pRS423 library Magnaporthe grisea cDNA clone p3fmgc_012733, mRNA sequence.
ACCESSION CK923106

```
VERSION     CK923106.1  GI:45391458
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   Dong,H., Peng,Y., Chen,B., Li,Y. and Li,D.
  TITLE     Large-scale identification of ESTs from Magnaporthe grisea by
            normalized cDNA library sequencing (2004b)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University, China Agricultural University, Guangxi
            University
            Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, china
            Tel: 0086-571-86961525
            Fax: 0086-571-86971183
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 reverse primer.
FEATURES             Location/Qualifiers
     source          1. .137
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /db_xref="taxon:148305"
                     /clone="p3fmgc_012733"
                     /tissue_type="Mycellium, conidium, germinating conidium,
                     swelling appressorium, mature appressorium, penetration
                     peg"
                     /dev_stage="Mycellium, conidium, germinating conidium,
                     swelling appressorium, mature appressorium, penetration
                     peg"
                     /clone_lib="Normalized Magnaporthe grisea cDNA pRS423
                     library"
                     /note="Vector: pRS423"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 7;  Length 137;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        126 AAGGTTATATAT 137


RESULT 15
BX327794
LOCUS       BX327794                146 bp    mRNA    linear   EST 07-APR-2004
DEFINITION  BX327794 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI026YN16 5-PRIME, mRNA sequence.
ACCESSION   BX327794
VERSION     BX327794.2  GI:46270728
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 2, 2003 this sequence version replaced gi:30336692.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            1798.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAG028ZH05_CS02632_1&c=1798.r
            .
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI026YN16"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 5;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          6 AAGGTTATATAT 17


RESULT 16
B95930/c
LOCUS       B95930                  147 bp    DNA     linear   GSS 31-MAR-1998
DEFINITION  T18I6TF TAMU Arabidopsis thaliana genomic clone T18I6, genomic
            survey sequence.
ACCESSION   B95930
VERSION     B95930.1  GI:2997868
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 147)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
            Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T18I6TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 147.
FEATURES             Location/Qualifiers
     source          1. .147
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T18I6"
                     /sex="hermaphrodite"
                     /clone_lib="TAMU"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 9;  Length 147;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
```

```
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         14 AAGGTTATATAT 3


RESULT 17
CA364088/c
LOCUS       CA364088                148 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  638929 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT72C17_A_B09 5',
            mRNA sequence.
ACCESSION   CA364088
VERSION     CA364088.1  GI:24672957
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 148)
  AUTHORS   Rexroad,C.E. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
            Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
  TITLE     Sequence analysis of a rainbow trout cDNA library and creation of a
            gene index
  JOURNAL   Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
COMMENT     Contact: Rexroad CE
            USDA, ARS, National Center for Cool and Cold Water Aquaculture
            11876 Leetown Road, Kearneysville, WV 25430, USA
            Tel: 304 724 8340 x2129
            Fax: 304 725 0351
            Email: crexroad@ncccwa.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329.
            Seq primer: AGCGGATAACAATTTCACACAGGA.
FEATURES             Location/Qualifiers
     source          1. .148
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="1RT72C17_A_B09"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="NCCCWA 1RT"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from brain, gill, liver,
                     spleen, muscle, and kidney."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 148;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        131 AAGGTTATATAT 120


RESULT 18
BG929548
LOCUS       BG929548                150 bp    mRNA    linear   EST 06-JUN-2001
DEFINITION  950006A12.x1 950 - Mature pollen from Sheila McCormick's lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BG929548
VERSION     BG929548.1  GI:14324071
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 950006  row: A  column: 12.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /tissue_type="pollen"
                     /dev_stage="mature"
                     /lab_host="SOLR"
                     /clone_lib="950 - Mature pollen from Sheila McCormick's
                     lab"
                     /note="Vector: Stratagene's Uni-Zap XR (pBluescript SK-);
                     Site_1: EcoRI (5-prime); Site_2: XhoI (3-prime);
                     Unamplified cDNA library directionally cloned by Rima
                     Kulikauskas using Stratagene's Uni-Zap system. Insert
                     sizes ranged from 0.5Kb to 2Kb.  50 microliter aliquot had
                     338,000 pfu when it was made in Sept, 1995,  from oligo
                     dT-primed poly A+ RNA."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         59 AAGGTTATATAT 70


RESULT 19
AV107977
LOCUS       AV107977                153 bp    mRNA    linear   EST 28-JUN-1999
DEFINITION  AV107977 Mus musculus liver C57BL/6J 13-day embryo Mus musculus
            cDNA clone 2510042D07, mRNA sequence.
ACCESSION   AV107977
VERSION     AV107977.1  GI:5255525
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 153)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
```

trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998)) Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998)) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .153
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2510042D07"
                     /sex="mixed"
                     /tissue_type="liver"
                     /dev_stage="13-day embryo"
                     /clone_lib="Mus musculus liver C57BL/6J 13-day embryo"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 1;  Length 153;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         52 AAGGTTATATAT 63


RESULT 20
BE058406/c
LOCUS       BE058406                155 bp    mRNA    linear   EST 14-JUL-2004
DEFINITION  sn15f09.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-11586 5', mRNA sequence.
ACCESSION   BE058406
VERSION     BE058406.1  GI:8402772
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 155)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 344   Std Error: 0.00
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-11586"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /clone_lib="Gm-c1016"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 155;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db         41 AAGGTTATATAT 30


RESULT 21
CA766087
LOCUS       CA766087                157 bp    mRNA    linear   EST 08-JAN-2003
DEFINITION  AF53-Rpf_10_A15_T7_050.ab1 IRRI Drought Stress Panicle Library
            Oryza sativa (indica cultivar-group) cDNA clone C0003471 5' similar
            to unknown, mRNA sequence.
ACCESSION   CA766087
VERSION     CA766087.2  GI:27548005
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 157)
  AUTHORS   Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and
            Bruskiewich,R.M.
  TITLE     IRRI Drought Stress Panicle cDNA Library
  JOURNAL   Unpublished (2002)
COMMENT     On Dec 2, 2002 this sequence version replaced gi:25995342.
            Contact: Richard Bruskiewich
            Biometrics and Bioinformatics Unit
            International Rice Research Institute
            DAPO 7777, Metro Manila, Philippines
            Tel: +63-2-845-0563
            Fax: +63-2-845-0606
            Email: r.bruskiewich@cgiar.org
            International Rice Information System (IRIS;
            http://www.iris.irri.org): D0203470
            Assignment of putative function to the sequence by S. Rudd of the
            Munich Information Center for Protein Sequences
            (http://mips.gsf.de)
            Plate: 10  row: A  column: 15.
FEATURES             Location/Qualifiers
     source          1. .157
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="C0003471"
                     /tissue_type="Panicles"
                     /dev_stage="Flowering"
                     /clone_lib="IRRI Drought Stress Panicle Library"
                     /note="Vector: pBluescript II SK+; Water stress was
                     applied by not watering for 4 consecutive days. Panicles
                     were collected from control (well watered) and stressed
                     plants at 2 days before heading, at heading, 50% flowering
```

and 4 days after 50% flowering."
ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 157;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 101 AAGGTTATATAT 112

RESULT 22
CO720276/c
LOCUS CO720276 158 bp mRNA linear EST 27-JUL-2004
DEFINITION tai41f12.y1 HyEch JUMY T1 Hydractinia echinata cDNA 5', mRNA sequence.
ACCESSION CO720276
VERSION CO720276.1 GI:50697149
KEYWORDS EST.
SOURCE Hydractinia echinata
ORGANISM Hydractinia echinata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae; Hydractiniidae; Hydractinia.
REFERENCE 1 (bases 1 to 158)
AUTHORS Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WashU Hydra EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Hans Bode
WashU Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Marcus Frohme and Uri Frank Library materials provided by Marcus Frohme, German Cancer Research Center(DKFZ) Heidelberg, Uri Frank, University of Heidelberg Library re-arrayed by Jorge Sozaried (DKFZ) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hrbode@uci.edu)
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .158
/organism="Hydractinia echinata"
/mol_type="mRNA"
/db_xref="taxon:35630"
/lab_host="DH10B"
/clone_lib="HyEch JUMY T1"
/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; A pool of mRNA was primed with an anchored oligo-dT adaptor with a Not I site for 1st strand synthesis. Double stranded cDNA was ligated to Sal-Adaptors and cut with NotI. After size selection (without radioactivity) cDNA was ligated into pSport with SalI/NotI termini. Transformation was in DH10b Electromax T1 phage resistant cells. Plating was on 2YT/Carbenicillin Agar and 14x384 cells were blue/white selected automatically picked into 2YT/HMFM/Carbenicillin (HMFM is a freezing additive)."
ORIGIN

Query Match 100.0%; Score 12; DB 7; Length 158;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 81 AAGGTTATATAT 70

RESULT 23
T26907/c
LOCUS T26907 159 bp mRNA linear EST 14-SEP-1998
DEFINITION ESTDIR575 CD34+DIRECTIONAL Homo sapiens cDNA clone CDDIR575 3', mRNA sequence.
ACCESSION T26907
VERSION T26907.1 GI:567798
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159)
AUTHORS Yang,Y., Peterson,K.R., Stamatoyannopoulos,G. and Papayannopoulou,T.
TITLE Human CD34+ cell EST database: single-pass sequencing of 402 clones from a directional cDNA library
JOURNAL Exp. Hematol. 24 (5), 605-612 (1996)
PUBMED 8605965
COMMENT Contact: Yi Yang
Medical Genetics, RG-25
University of Washington
Health Science Building, Room BB512, Seattle, WA 98195
Tel: 2065432368
Fax: 2065433050
Email: yi629@u.washington.edu
Seq primer: T7 promoter.
FEATURES Location/Qualifiers
source 1. .159
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CDDIR575"
/clone_lib="CD34+DIRECTIONAL"
/note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human CD34+ cells. cDNA was synthesized using Stratagene's ZAP-cDNA synthesis kit(Cat#200400). cDNA was directionally cloned into EcoRI and XhoI sites in vector Lambda ZAPII"
ORIGIN

Query Match 100.0%; Score 12; DB 8; Length 159;
Best Local Similarity 100.0%; Pred. No. 2.5e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTTATATAT 12
||||||||||||
Db 29 AAGGTTATATAT 18

RESULT 24
CD162149
LOCUS CD162149 160 bp mRNA linear EST 14-SEP-2003
DEFINITION ML1-0075T-R173-D01-U.G ML1-0075 Schistosoma mansoni cDNA clone ML1-0075T-R173-D01.G, mRNA sequence.
ACCESSION CD162149
VERSION CD162149.1 GI:34698864
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 160)
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,

```
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0075T-R173  row: 1  column: D.
FEATURES             Location/Qualifiers
     source          1. .160
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0075T-R173-D01.G"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0075"
                     /note="Vector: pGEM T-easy"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 6;  Length 160;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        132 AAGGTTATATAT 143


RESULT 25
BG513953/c
LOCUS       BG513953                161 bp    mRNA    linear   EST 28-MAR-2001
DEFINITION  dac15h11.y2 NICHD_XGC_He1 Xenopus laevis cDNA clone IMAGE:4407404
            5', mRNA sequence.
ACCESSION   BG513953
VERSION     BG513953.1  GI:13484610
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 161)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L.,
            Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y.,
            Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .161
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4407404"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_He1"
                     /note="Organ: heart; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 161;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        126 AAGGTTATATAT 115


RESULT 26
CE584959
LOCUS       CE584959                163 bp    DNA     linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-17000366332200 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE584959
VERSION     CE584959.1  GI:36901740
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db          7 AAGGTTATATAT 18


RESULT 27
BX655471/c
LOCUS       BX655471                165 bp    DNA     linear   GSS 04-APR-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-589E01-021428,
```

```
            genomic survey sequence.
ACCESSION   BX655471
VERSION     BX655471.1  GI:37611859
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
  TITLE     GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for
            the identification of T-DNA insertion mutants in Arabidopsis
            thaliana
  JOURNAL   Bioinformatics 19 (11), 1441-1442 (2003)
   PUBMED   12874060
REFERENCE   2
  AUTHORS   Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and
            Weisshaar,B.
  TITLE     An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
            flanking sequence tag-based reverse genetics
  JOURNAL   Plant Mol. Biol. 53 (1-2), 247-259 (2003)
   PUBMED   14756321
REFERENCE   3
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and
            Weisshaar,B.
  TITLE     High-throughput generation of sequence indexes from T-DNA
            mutagenized Arabidopsis thaliana lines
  JOURNAL   BioTechniques 35 (6), 1164-1168 (2003)
   PUBMED   14682050
REFERENCE   4  (bases 1 to 165)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence has been recovered from the left border of the T-DNA.
            It indicates an insertion within the locus defined by BAC clone
            fca_all. Details on the protocols used for generation of the
            sequence are described in References 1-3. The sequences are
            generated at the MPI for Plant Breeding Research in the context of
            the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
            program designated 'GABI'. Information on line availability can be
            found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .165
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="GK-589E01-021428"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Col-0"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161 (GenBank accession number: AJ537514). The
                     lines contain one or more T-DNA insertions. The DNA
                     fragment(s) resulting from the PCR were directly sequenced
                     to determine the genomic sequence flanking the insertion.
                     T-DNA derived sequences were removed."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 10;  Length 165;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        129 AAGGTTATATAT 118


RESULT 28
CV314160/c
LOCUS       CV314160                166 bp    mRNA    linear   EST 24-SEP-2004
DEFINITION  CM1-FN0103-030700-296-b01 FN0103 Homo sapiens cDNA, mRNA sequence.
ACCESSION   CV314160
VERSION     CV314160.1  GI:52637374
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 166)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. http://www.ludwig.org.br.
FEATURES             Location/Qualifiers
     source          1. .166
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="FN0103"
                     /note="Organ: prostate_normal; Vector: puc18; Site_1:
                     SmaI; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 7;  Length 166;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        155 AAGGTTATATAT 144


RESULT 29
BF911105
LOCUS       BF911105                168 bp    mRNA    linear   EST 18-JAN-2001
DEFINITION  PM1-UT0061-061100-010-a06 UT0061 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF911105
VERSION     BF911105.1  GI:12302563
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 168)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
```

```
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-UT0061-
            061100-010-a06&t3=2000-11-06&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 135.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UT0061"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             100.0%;  Score 12;  DB 2;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        132 AAGGTTATATAT 143


RESULT 30
BP616314/c
LOCUS       BP616314                 168 bp    mRNA    linear   EST 26-JUN-2004
DEFINITION  BP616314 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-19-I08 3',
            mRNA sequence.
ACCESSION   BP616314
VERSION     BP616314.1  GI:49267496
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 168)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-19-I08"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN

  Query Match             100.0%;  Score 12;  DB 3;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+04;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGTTATATAT 12
              ||||||||||||
Db        114 AAGGTTATATAT 103


Search completed: December 20, 2005, 15:08:08
Job time : 63.7751 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 21, 2005, 10:28:07 ; Search time 3.01606 Seconds
(without alignments)
5614.164 Million cell updates/sec

Title: US-09-837-344-46
Perfect score: 20
Sequence: 1 AAGGTTATATAT 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4324718

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09837344/runat_21122005_093215_22916/app_query.fasta_1.3740
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09837344_@CGN_1_1_1319_@runat_21122005_093215_22916 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description

No matches found

Search completed: December 21, 2005, 11:27:08
Job time : 3.01606 secs

THIS PAGE BLANK (USPTO)